Master Thesis

Converting Near Infrared Facial Images to Visible Light Images using Skin Pigment Model

皮膚色素モデルに基づく近赤外顔画像から可視光顔画像への変換

Dept. of Information & Communications Engineering
Grad. Sch. of Information Science and Technology
The University of Tokyo

48–116421

Goh Kim Shing

Advisor Professor Yoichi Sato

February 2013

Converting Near Infrared Facial Images to Visible Light Images
using Skin Pigment Model

by

Goh Kim Shing

Abstract

Facial images are the primary biometric used for a human identification system and a tracking system. Especially, a visible light (VIS) facial image is commonly used as it is easy to be acquired and its appearance is similar to the human visual system. However, images taken under varying illumination or a low-lit condition result in poor recognition accuracy for systems using VIS images. To alleviate this problem, a system using an active near infrared (NIR) flash is proposed to acquire illumination invariant facial images. Yet, this system requires both probe and gallery images to be taken under NIR spectrum which limits the applicability of the system as NIR images are often unavailable for gallery images, *e.g.*, photos on a passport or a driving license. In such cases, probe NIR images are matched against VIS gallery images. One of the solutions to this problem is to convert probe NIR images to artificial VIS images before matching.

In this thesis, we propose a physics-based method to synthesize artificial facial images in visible wavelengths from multiband NIR images. Studies on photometric properties of human skin show that melanin and hemoglobin components are dominant factors that affect the skin appearance under different light spectrum. Our physics-based synthesis method takes account of the photometric properties of human skin, introducing a skin pigment model. Specifically, a set of intensities observed at a certain surface point with varying wavelength is represented by a linear combination of both the pigment components. Our method learns the spectral basis vectors, which describe absorbance of both the pigments, from multispectral image dataset by using Independent Component Analysis (ICA). Then, our method estimates the coefficients, which are pixel-wise densities of both the pigments, from a multiband NIR image, and finally converts it to a visible light image. We demonstrate that our method works well for real facial images even though only a small dataset is available for learning basis vectors.

Acknowledgement

I would like to express my gratitude to all those who gave me the possibility to complete this thesis.

First, I would like to express my deepest appreciation to my research advisor, Professor Yoichi Sato for his kind guidance, strong support, and providing a good environment for my research with helpful staffs allowing me to pursue my research interest. Without his guidance and persistent help this thesis would not have been possible.

Next, I would like to thank Mr. Takahiro Okabe and Mr. Tetsu Matsukawa who have given me instructions about the research and have taught me a lot about the techniques of experiments.

I also appreciate Mr. Yusuke Sugano for sharing his knowledge in research and thesis writing skills with me. He has also cheered me a lot in my whole research life in Sato Laboratory.

At the same time, I wish to express my thanks to Sato Laboratory staffs and members. In particular, Keisuke Ogaki, Yasuyuki Kobashi, Wiennat Mongkulmann, who are my colleagues who helped me a lot in my research and my life in this laboratory. And also, special thanks goes to all members in Sato Laboratory have made my two year studies in the University of Tokyo a wonderful times.

Last but not least, I wish to thanks my parents, all of my family members and friends, for their persistent encouragement and assistance.

Table of Contents

List of Figures and Tables

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation of Research

Traditionally, passwords and ID cards have been widely used to restrict access but these methods can be easily violated and are unreliable. Biometrics which based on one's body characteristics, *e.g.*, face, fingerprint, or iris, are unique and cannot be borrowed, forgotten or stolen [1]. Therefore, in recent years, biometrics has received significant attentions for personal authentication systems due to its reliability compared to traditional methods.

Among all the biometrics, facial images are the most common biometric characteristic used to make a personal recognition due to its availability and high compatibility with auxiliary human visual recognition. For this reason, International Civil Aviation Organization (ICAO) has identified face images as the primary biometrics with fingerprints and iris recognition as backup for use in Machine Readable Travel Documents[1], such as ePassport, in early 2001.

Face recognition involves feature set extraction from a two-dimensional probe image of the user's face and matching it with the gallery images stored in a database. Although researches on face recognition have reached a certain level of maturity, their recognition accuracy in real applications are still limited by various condition. VIS images are commonly used due to its generality. One of the major challenges in VIS face recognition is how to deal with images taken under varying illumination or a low-lit condition [2]. To alleviate the problem, Li *et al.* [3] proposed a system using active near infrared flash to acquire illumination invariant facial images. However, their system is based on the strong assumption that NIR images are available as gallery images as well as probe images. This limits the applicability of the system because NIR images are often unavailable for gallery images, *e.g.*, photos on a passport or a driving license. In such cases, probe NIR images are matched against VIS gallery images, resulting in poor recognition accuracy. This is because appearance of a face changes significantly between NIR and VIS images as shown in Figure 1.1.

Several methods for matching NIR images to VIS images have been proposed. One approach focuses on extracting features invariant, *e.g.*, local binary pattern (LBP) [4] and histograms of oriented gradients (HOG), across both NIR and VIS spectra [5, 6]. Another approach tries to

[1] http://www.icao.int/publications/Documents/9303_p3_v1_cons_en.pdf

Figure 1.1: Examples of NIR and corresponding VIS images.

synthesize a VIS image from a NIR image [7–10]. The advantage of the latter approach is that existing face recognition systems can be used with no modification. Moreover, the VIS image synthesized from NIR image can be used for auxiliary human visual recognition when current system fail to function, where hybrid human-computer recognition can be implemented. For these reasons, research on face synthesis from NIR and VIS has recently been paid more attentions.

Chen *et al.* [8] proposed a method for NIR to VIS image conversion based on local linear embedding. Each image patch is approximated by a weighted sum of their k-nearest neighbors (KNN) of training NIR patches using local binary pattern (LBP) similarities. Then a VIS image is synthesized by using the same weights and corresponding VIS patches. Zhang *et al.* [9] extended this idea by using sparse representation. Shao *et al.* [10] learned the relationship of VIS and NIR images by using a multifactor analysis. Similarly, Zhang *et al.* [7] learned the relationship of NIR and quotient images. All of these methods share the same problem that a large number of patch pairs are necessary to cover various face appearances, *e.g.*, patches collected from more than 100 individuals, to produce satisfactory results. This is because these methods try to convert NIR patches to VIS patches without considering the underlying physical phenomenon.

1.2 Purpose of Research

In this thesis, we focus on the face synthesis approach for NIR-VIS matching, which in particular, we propose a method for converting NIR images into VIS images based on the photometric properties of human skin.

Human skin is a multi-layered structure with various pigments. Melanin and hemoglobin pigment are dominant pigments that affect skin appearance [11, 12]. Tsumura *et al.* [13] proposed a technique to extract melanin and hemoglobin bases and densities from a RGB spectral by using ICA [14]. They synthesized various skin colors such as tanning and alcohol consumption by changing the extracted pigment densities. However, their analysis was limited to RGB color channels, and no discussion was made for skin appearances in the NIR spectrum.

In this paper, we propose a method for converting NIR images to a VIS images on the basis of the skin pigment model. Our method estimates the coefficients, which are pixel-wise densities of two dominant pigments, from a multiband NIR image, so at least two images under different spectra in NIR[1] is needed. We can synthesis the images of VIS spectral from the estimated pigment density and spectral basis.

As far as we know, this work is the first attempt to extract skin pigment density from multispectral NIR images and synthesize VIS images by using skin pigment model. Our experiments demonstrate that the proposed method is able to synthesize VIS images without requiring a large number of training samples.

1.3 Thesis Overview

The rest of this thesis is organized as follows. In Chapter 2, some related works on heterogeneous face image synthesis are introduced. In Chapter 3, optical properties of human skin and research on skin reflectance analysis are reviewed. Then we introduce the proposed method in Chapter 4, followed with our experimental results in Chapter 5 Lastly, in Chapter 6, we conclude with a summary on our works followed by some discussions on future works.

[1]The use of special equipment for capturing multispectral NIR images could be a limitations of the proposed method from a practical point of view. Recently, however, such equipment is getting more popular in the field of multispectral imaging [15, 16], and could be used for our purpose.

Chapter 2

Related Research for Heterogeneous Face Synthesis

Face images taken with different modalities are heterogeneous because they have different intrinsic characteristics. Due to this difference, heterogeneous face images have distinct difference in appearance and cannot be direct matching results in poor recognition accuracy. One of the approaches is to transform the heterogeneous face images to homogeneous face images by performing heterogeneous synthesis before matching. Heterogeneous synthesis is a process of generating a face image from an input which is taken from other modalities, given a set of sample taken with those 2 modalities. For example, synthesizing a VIS image from a sketch, or a depth (3D) image, or a NIR image and vice versa.

This chapter focuses on introducing the researches related to heterogeneous face synthesis. First section introduces face sketch synthesis method. Second section introduces works on depth image prediction from VIS image which is also applicable to NIR image synthesis. Third section introduces research on face synthesis from NIR to VIS which is also our main concern in this thesis. In the last section, summary of related works is discussed.

2.1 Face Sketch Synthesis

One of the interesting works in heterogeneous synthesis is face sketch synthesis. We introduce the face sketch synthesis for a comprehensive understanding of image conversion from one modality to the other. The learning based frameworks in face sketch synthesis are similar to other NIR-VIS synthesis method to be introduced later in this chapter.

Tang and Wang [17] proposed a Principal Component Analysis (PCA) based method for face sketch synthesis. Figure 2.1 shows the framework of their proposed synthesis method.

First, all fiducial points on an input face image P, are located to extract the shape information. Next, the face image is transformed by image warping to a mean face shape derived from training set using face graph matching [18] to separate the texture I_p and shape G_p from the original image. Then, eigentransformation [19] is applied to I_p and G_p respectively to generate texture I_s and shape G_s for the synthesized sketch image before warping them back to produce final synthesized sketch image S. In eigentransformation, the transformation between original

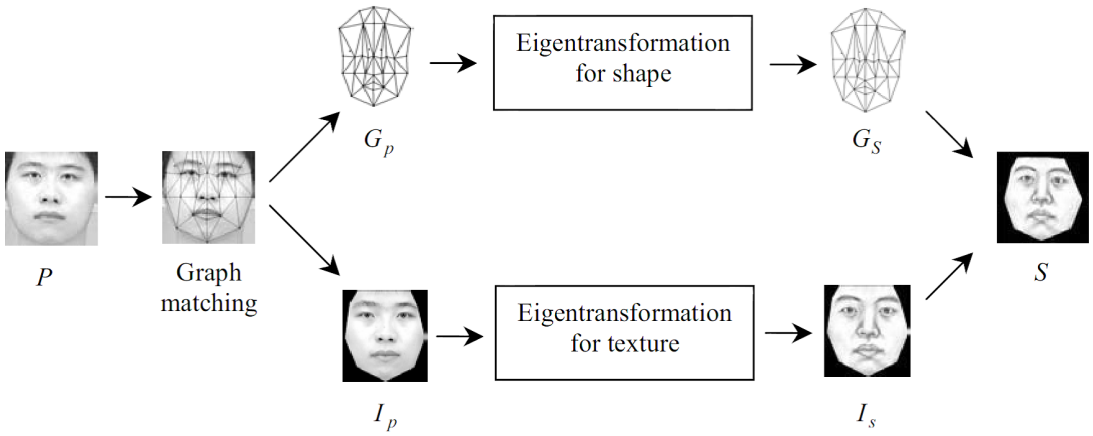

Figure 2.1: Framework of Tang and Wang's method for face sketch synthesis [17].

image and sketch is assumed to be approximated as a linear process. They approximately represent the input image P as weighted linear combination of K training images as shown in Equation (2.1):

$$\vec{P} = \sum_{i=1}^{K} c_i(\vec{P}_i - \vec{m}_p) + \vec{m}_p, \tag{2.1}$$

where c_i is weight coefficient for the i-th training image decided by PCA, $\vec{P}_i$ is the i-th training image expressed as a vector, and $\vec{m}_p$ is a mean vector for all training images. Replacing each training image P_i with its corresponding sketch S_i, a synthesized sketch S can be reconstruct as follows:

$$\vec{S} = \sum_{i=1}^{K} c_i(\vec{S}_i - \vec{m}_s) + \vec{m}_s, \tag{2.2}$$

where $\vec{m}_s$ is a mean vector for sketch corresponding to all training images.

Figure 2.2 shows some results of face sketch synthesis using Tang and Wang's method. The results shows that sketch textures are transformed and similar to the real sketches drawn by pencil. However, their method works only for frontal face images taken under a controlled condition with not many changes in expression, view, and illumination.

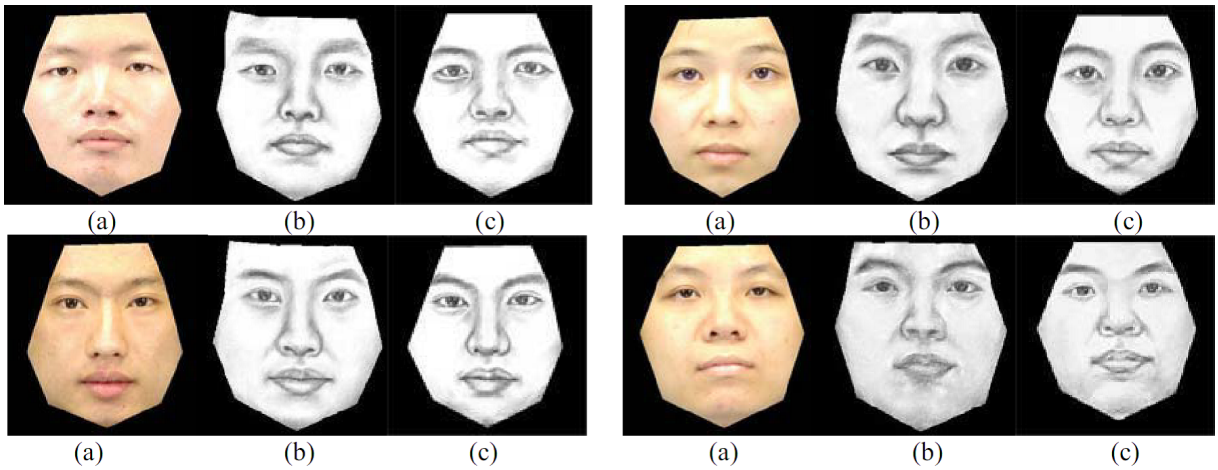

Figure 2.2: Face synthesis result using Tang and Wang's method: (a) original face image; (b) synthesized sketches using Tang and Wang's method; (c) real sketches drawn by artist [17].

Meanwhile, Liu *et al.* [20] employ a patch-based strategy for the face sketch synthesis task. Their proposed method is based on the idea of local geometry preserving for pseudo-sketch synthesis, given a set of training photo and sketch image pairs, T_p and T_s. Subscripted p and s indicate "photo" and "sketch", respectively. They divide the photo and sketch images into N patches in an overlapped manner and denote each photo and sketch patches as $\hat{I}_p^t$ and $\hat{I}_s^t$ where $t = 1, 2, ..., N$.

For an input photo, the image is divided into patches using the same manner as the training images. For each patches I_p^t, the Euclidean distance is used to find its KNN $\hat{I}_{pk}^t$, $k = 1, 2, ...K$ in T_p. Next, the weights w_{pk}^t of the neighbors are computed by minimizing the errors in reconstructed I_p^t. Each pseudo-sketch patch I_s^t is then estimated using the corresponding sketch patches $\hat{I}_{sk}^t$ of KNN $\hat{I}_{pk}^t$ and the weights w_{pk}^t, $k = 1, 2, ..., K$ as follows:

$$I_s^t = \sum_{k=1}^{K} w_{pk}^t \hat{I}_{sk}^t. \tag{2.3}$$

Lastly, all patches are assembled and overlapped region are averaged in final result.

Their sketch synthesis results are shown in Figure 2.3 with comparison to the synthesis results using eigentransform method. Their results can well approximate the real sketches while the eigentransform method fails to reconstruct some local facial information. This is because their method uses local geometry preserving to approximate the complex nonlinear mapping between photo and sketch patches, while the eigentransform method simply solves it as a linear process. However, some artifacts such as blurs still exist in their result because of the weighted sum of KNN patch and overlapped region.

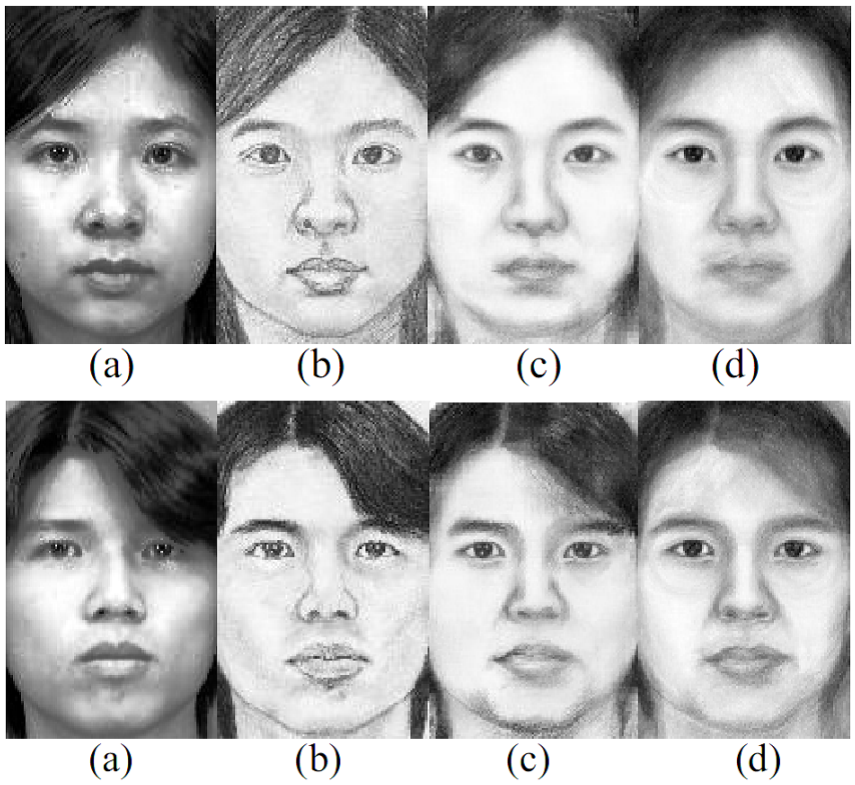

Figure 2.3: Face sketch synthesis results using Liu's method: (a) original photo image; (b) sketch drawn by artist; (c) sketch synthesized using Liu's method; (d) synthesized sketch using eigentransform method [20].

2.2 3D Facial Image Synthesis

Reiter *et al.* [21] apply canonical correlation analysis (CCA) and coupled statistical model (CSM) respectively to recover 3D face shape or NIR image from a two-dimensional RGB image. CCA is suitable for building relationship between two set of data or signals, for example, in Reiter's work, between 3D data/NIR image and RGB image. Here, we focus on the discussion on NIR prediction from RGB image as it is more related to our goal. In their work, NIR and RGB patches are obtained by employing active appearance models [22].

Given N pairs of RGB patches $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N)$ and corresponding NIR patches $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_N)$, Reiter *et al.* employed CCA to find the leading factor pairs $\boldsymbol{W} = (\boldsymbol{w}^1, \boldsymbol{w}^2, ..., \boldsymbol{w}^k)$ and $\boldsymbol{V} = (\boldsymbol{v}^1, \boldsymbol{v}^2, ..., \boldsymbol{v}^k)$ where k is the lower value of dimensionality space of RGB image and NIR image. Both $\boldsymbol{W}$ and $\boldsymbol{V}$ project $\boldsymbol{X}$ and $\boldsymbol{Y}$ in such that $\boldsymbol{W}^T\boldsymbol{X}$ and $\boldsymbol{V}^T\boldsymbol{Y}$ are most correlated. For a new input $\boldsymbol{x}_{new}$, the predicted $\boldsymbol{y}_{new}$ is computed as follows:

$$\boldsymbol{y}_{new} = \boldsymbol{R}^T\boldsymbol{p}, \tag{2.4}$$

where $\boldsymbol{R} = (\boldsymbol{X}^T\boldsymbol{W})^\dagger\boldsymbol{Y}^T$ and $\boldsymbol{p} = \boldsymbol{W}^T\boldsymbol{x}_{new}$.

Figure 2.4 shows their results for NIR reconstructions. The results show only small errors for NIR predictions. However, in their methods, both the RGB and NIR images are projected into vectors, which ignore the intrinsic structure of image and the derived vectors is usually of high dimension. These are likely to cause the curse of dimensionality when there are only limited training data.

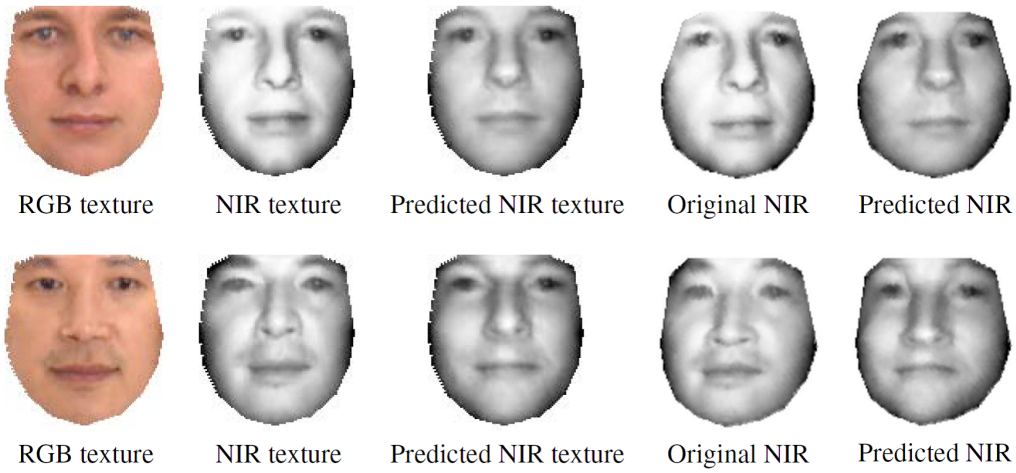

Figure 2.4: NIR prediction using Reiter's method [21].

2.3 VIS Facial Image Synthesis

Chen *et al.* [8] proposed a face synthesis method to convert a NIR face image into VIS domain. Figure 2.5 shows Chen's method composed of two procedures which are training procedure and testing procedure.

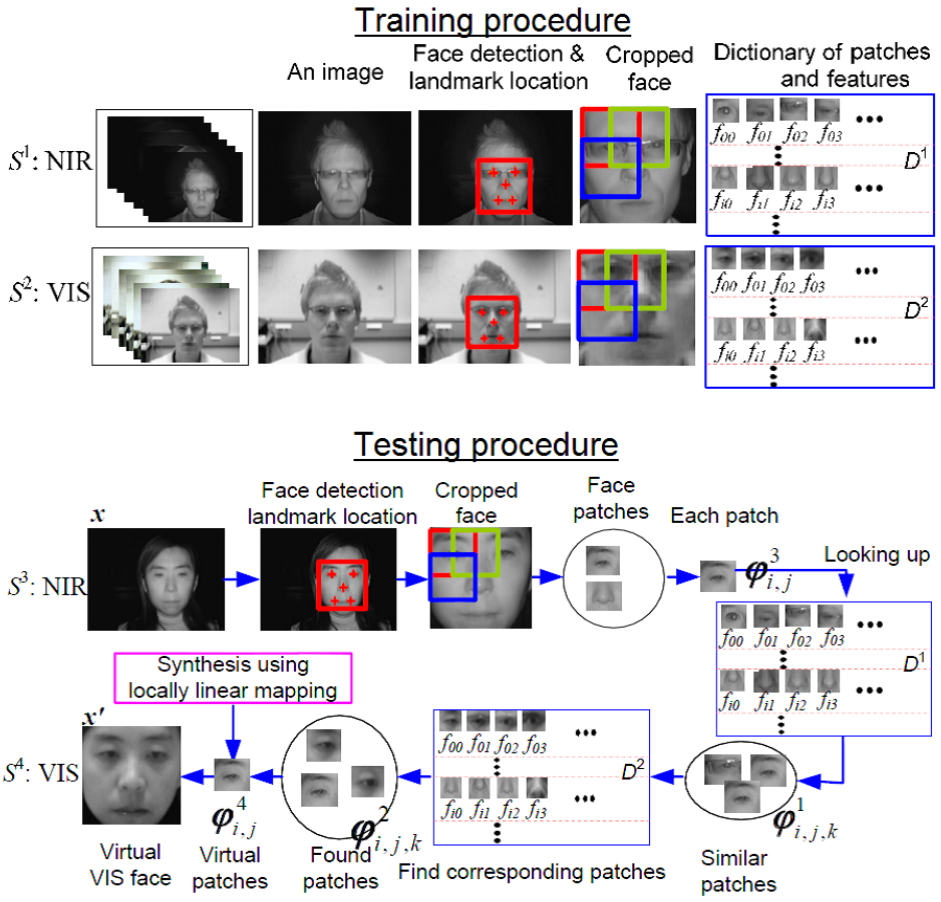

Figure 2.5: Training procedure and testing procedure of Chen's method [8].

In training procedure, for all NIR-VIS image pairs in S^1 and S^2, location of two eyes, nose and two mouth corners are detected. Face image are cropped and normalized to a fixed size based on five locations detected above. Next, all images are divided into patches in an overlapped way. Each patch is denoted as $\boldsymbol{\varphi}_{i,j}$, where i and j are the indexes of patches in the row and column direction, respectively. For each patch, feature $\boldsymbol{f}_{i,j}$ is computed and combined with its patch $\boldsymbol{\varphi}_{i,j}$ to obtain a dictionary $\boldsymbol{D}^1$ and $\boldsymbol{D}^2$ for each training set S^1 and S^2. They use multi-resolution local binary pattern (MLBP) to describe feature for each patch.

In testing procedure, for an input image $\boldsymbol{x}$ in S^3, same process is perform on $\boldsymbol{x}$ to obtain the feature $\boldsymbol{f}^3_{i,j}$ for each patch $\boldsymbol{\varphi}^3_{i,j}$. Next, for each input patch $\boldsymbol{\varphi}^3_{i,j}$, they look up dictionary $\boldsymbol{D}^1$ for its KNN $\boldsymbol{\varphi}^1_{i,j,k}$, $k = 1, 2, ...K$. Replace each KNN $\boldsymbol{\varphi}^1_{i,j,k}$ with its corresponding patch $\boldsymbol{\varphi}^2_{i,j,k}$ in

$\boldsymbol{D}^2$, they compute the synthesized patch as follow:

$$\boldsymbol{\varphi}^4_{i,j} = \sum_{k=0}^{K-1} \omega_k \boldsymbol{\varphi}^2_{i,j,k}, \tag{2.5}$$

where ω_k are the normalized weights determined using histogram intersection as a similarity of two MLBP histograms. Lastly, all synthesized patches are combined to obtain final result $\boldsymbol{x}'$. For overlapped regions, average values are used for output.

Figure 2.6 shows synthesized results using Chen's method. Their synthesized results look fine except the first one. For the first one, the synthesized VIS image shows different expression to the input NIR image. This is because their samples in training set are insufficient to cover variation of expressions.

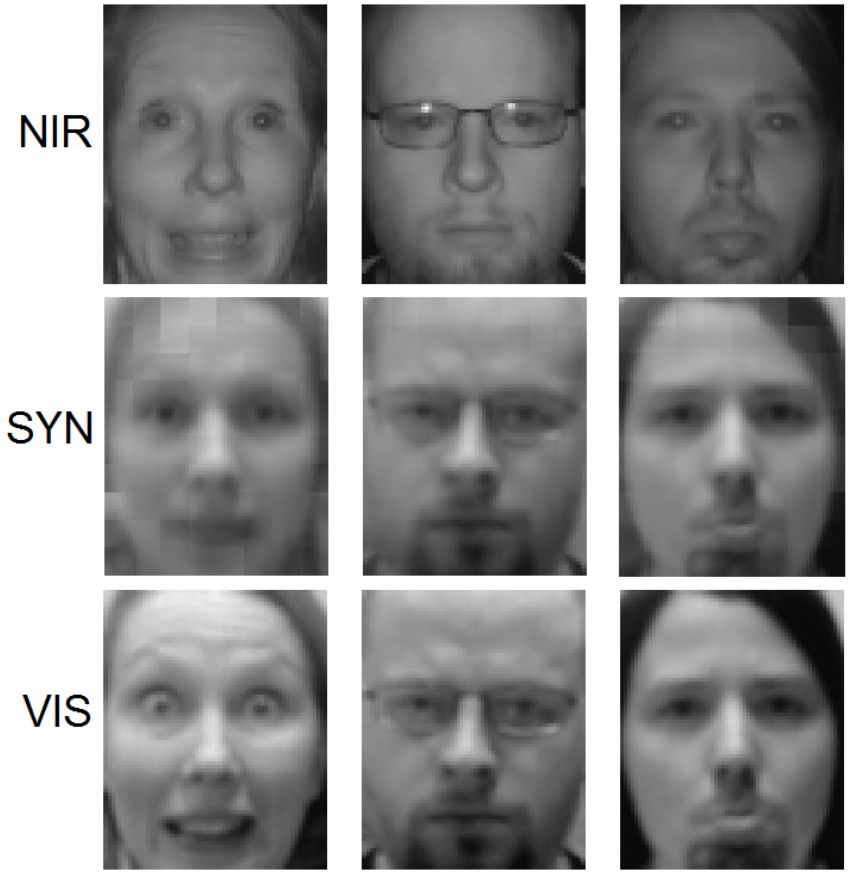

Figure 2.6: Face synthesis result using Chen's method [8].

Zhang *et al.* [7] propose a framework for synthesizing an artificial VIS face image from NIR input by taking account of the photometric properties of human skin. They introduce the use of quotient images [23, 24] for training and reconstruction procedure. According to Lambertian reflectance model, images captured under VIS and NIR condition can be written as follows:

$$I_{vis}(x,y) = \rho_{vis}(x,y)\vec{n}_{vis}(x,y) \cdot \vec{l}_{vis}(x,y), \tag{2.6}$$

$$I_{nir}(x,y) = \rho_{nir}(x,y)\vec{n}_{nir}(x,y) \cdot \vec{l}_{nir}(x,y), \tag{2.7}$$

where $I(x,y)$ refers to intensity of the captured image, $\rho(x,y)$ refers to albedo of skin surface, $\vec{n}(x,y)$ refers to normal vector of skin surface, and $\vec{l}(x,y)$ refers to direction and intensity of light source at point(x,y). By assuming that $\vec{n}_{vis}(x,y) \approx \vec{n}_{nir}(x,y)$ and $\vec{l}_{vis}(x,y) \approx \alpha\vec{l}_{nir}(x,y)$ where α is a constant scalar that indicates the ratio of lighting intensity for each spectrum, quotient image can be calculate by dividing Equation (2.6) with Equation (2.7), obtaining

$$Q(x,y) \approx \alpha\frac{\rho_{vis}(x,y)}{\rho_{nir}(x,y)}. \tag{2.8}$$

Note that in controlled lighting condition, α can be eliminated by normalizing lighting intensity. Figure 2.7 shows an example of quotient image obtained after illumination normalization. Normalized quotient image contains only information of albedo. By introducing the use of quotient image, they put only the intrinsic part, which is albedo ρ into the training process and leave out the $\vec{n} \cdot \vec{l}$ part. They claim that use of quotient image instead of NIR image in training process helps to preserve more detail information.

Figure 2.7: Example of normalized quotient image. (a)VIS image, (b)NIR image, (c)quotient image after illumination normalization [7].

Figure 2.8 shows the framework of Zhang's method. They employ tensorface [25] as a multilinear analysis tool to handle image ensembles that involve multiple factors, such as spectral channels and subject identities. Higher-order singular value decomposition (HOSVD) is applied to incorporate influence of each factor properly. They extend the idea of tensorface via kernel-based strategy [26], translating all facial feature to a high dimensional feature space F. For image patch at each location, instead of the original feature space, a tensor space F is built by a kernel-defined nonlinear transformation to satisfy their mapping relationship. Synthesized patches are expressed in form of a weighted sum of training images and overlapped regions are averaged.

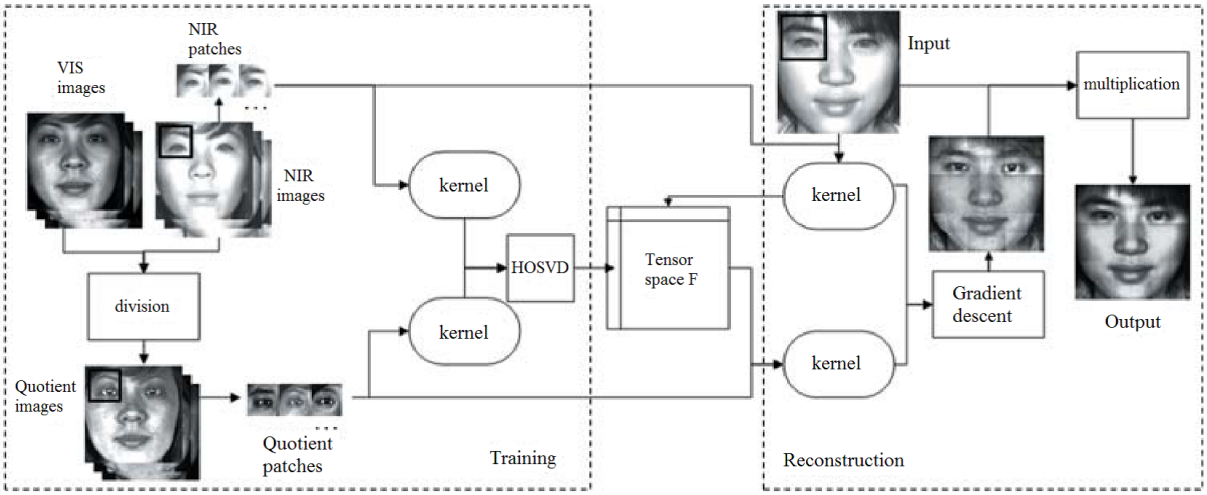

Figure 2.8: Framework of Zhang's method [7].

Figure 2.9 shows some results of Zhang's method with comparisons to other methods. They compared their indirect training method (training with quotient image and VIS pairs) to direct learning method (training with NIR and VIS pairs). By comparing (c) to (b) and (e) to (d), one can see that results with indirect learning are better. By comparing (e) to (d), synthesized images using kernel-based methods appear to be more natural to ground truth. However, some artifacts still exist in their synthesized result because KNN-based approach is dependent on the training data and would easily overfitting.

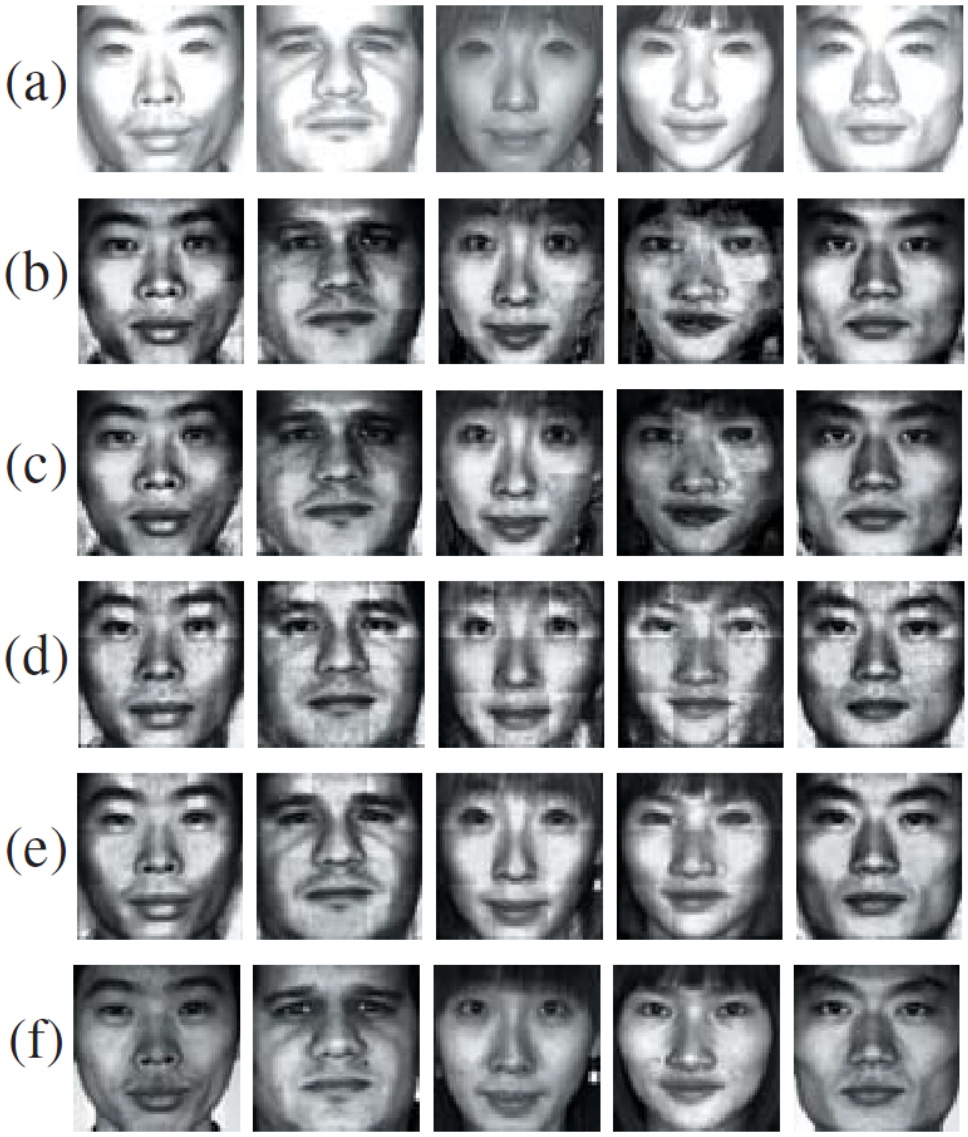

Figure 2.9: Comparison of (a)NIR input images, (b)direct multilinear approach, (c)indirect multilinear approach, (d)direct learning with kernel-based extension, (e)indirect learning with kernel-based approach, (f)ground truth VIS images [7].

Shao *et al.* [10] propose a super-resolution-based VIS face synthesis method. Their main idea is to enhance the quality of NIR image using tensorface [25], super resolution [27] and image fusion [28] techniques. Figure 2.10 shows their framework of proposed method composed of 3 steps which are building tensor spaces, super-resolution on tensor space and image fusion. First, a tensor containing different identities and spectral is constructed in high-resolution and low-resolution space for training purpose. Next, in testing phase, the input NIR image is projected into low-resolution tensor space, and its identity vector is super-resolved for high-resolution face image in VIS domain. Lastly, the simulated image of high-resolution is fused with original NIR image for better detail information using image fusion technique. They employ three methods in image fusion step, which are pixel-by-pixel averaging, PCA fusion and wavelet based fusion.

Figure 2.11 shows the synthesized result of Shao's method. Simulated VIS images after super-resolution step are shown in (b). Simulated VIS images are rich in texture which can be seen under VIS condition. (e) shows the results of their proposed method which look better compared to other fusion methods.

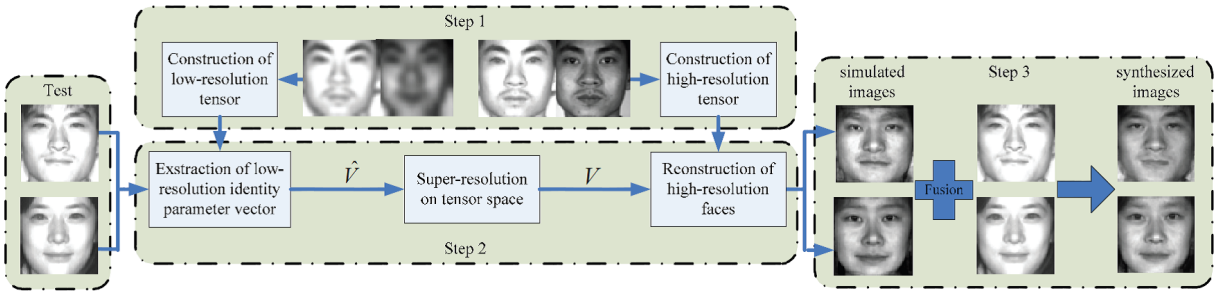

Figure 2.10: Training procedure and testing procedure of Shao's method [10].

Figure 2.11: Face synthesis result using Shao's method [10].

Wang *et al.* propose a pixel-wise synthesis method, called "face analogy", for converting face images from NIR to VIS. Face analogy shares the idea of "image analogy" [29]. Image analogy is an idea of transforming an image A to image A' in the same way as the given image B relates to given image B'.

They introduce a local normalization method to construct common local invariants of NIR and VIS image. Taking I_{vis} and I_{nir} as VIS face image and NIR face image. At position i, both image can be normalized as:

$$h_{vis}(i) \leftarrow \frac{I_{vis}(i) - m_{vis}(\aleph_i)}{\sigma_{vis}(\aleph_i)}, \tag{2.9}$$

$$h_{nir}(i) \leftarrow \frac{I_{nir}(i) - m_{nir}(\aleph_i)}{\sigma_{nir}(\aleph_i)}, \tag{2.10}$$

where $m(\cdot)$ and $\sigma(\cdot)$ is the mean and deviation of a region in a small neighborhood $\aleph_i$ of pixel i of face images in NIR and VIS domain, respcetively. Assume that there is a local linear transformation to equate $h_{vis} = h_{nir}$, the synthesized VIS image can be calculated as Equation2.11

$$I_{nir}(i) = \frac{I_{nir}(i) - m_{nir}(\aleph_i)}{\sigma_{nir}(\aleph_i)} \cdot \sigma_{vis}(\aleph_i) + \aleph_{nir}(\aleph_i) \tag{2.11}$$

Figure 2.12 shows synthesized results of Wang's method. Their results are reasonable and with less artifacts due to the pixel-wise transformation. However, we can see that the eye regions are not well-synthesized.

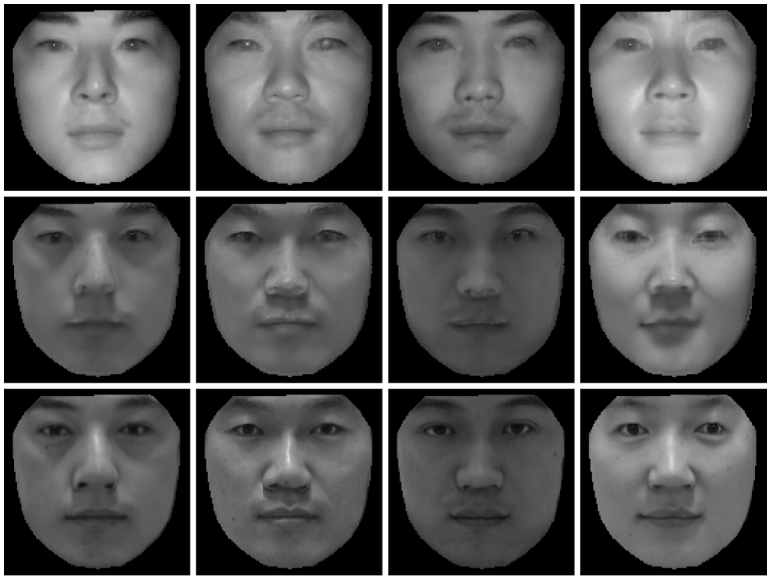

Figure 2.12: Face synthesis result using Wang's method. NIR inputs (first row); synthesized VIS (second row); ground truth VIS (third row) [30].

2.4 Summary of Related Research

In this chapter, we have introduced several related works on heterogeneous face synthesis. Most of the heterogeneous face synthesis methods are based on local linear embedding (LLE) approach [31]. In LLE approach, each patch is represented as a weighted sum of KNN in space of modalities, respectively. Weights estimated from one modality are transferred into another modality for reconstruction. For methods using LLE approach [7, 8, 20], which are also patch-based synthesis approach, issue of artifacts in synthesized result are unavoidable. Parameters such as, number of neighbors K, and patch size, need to be optimized for optimum performance. Large value of K is the cause for blur in synthesized result. On the other hand, big patch size results in noise while small patch size causes the result to lose some details. Patch-based approach requires a large amount of training samples to cover the variation of skin appearance. Shao *et al.* [10] manage to reduce the size of training set by employing super-resolution approach in their method. However, their training set still includes images from 50 subjects. Table 2.1 shows the comparison on size of training set between related works.

Table 2.1: Comparison on size of training set between related works.

	Input images	Output images	Size of training set
Tang & Wang [17]	Photo (VIS)	Sketch	306 samples from 306 subjects
Liu *et al.* [20]	Photo (VIS)	Sketch	306 samples from 306 subjects
Reiter *et al.* [21]	VIS	Depth (3D)	150 samples
		NIR	100 samples
Chen *et al.* [8]	NIR	VIS	1500 samples from 250 subjects
Zhang *et al.* [7]	NIR	VIS	99 subjects
Shao *et al.* [10]	NIR	VIS	50 samples from 50 subjects
Wang *et al.* [30]	NIR	VIS	198 samples from 99 subjects

For related research on synthesizing VIS image from NIR input, we summarize the related research based on their approach as shown in Table 2.2

Table 2.2: Summary of related research on synthesizing VIS image from NIR input based on their synthesis scales and techniques used.

	Scale	Approach & Techniques
Chen *et al.* [8]	Patch-based	LLE
Zhang *et al.* [7]	Patch-based	LLE, training with quotient images
Shao *et al.* [10]	Pixel-wise (Super-resoluted)	Tensorface, super-resolution, image fusion
Wang *et al.* [30]	Pixel-wise	Face analogy

Chapter 3

Optical Properties of Human Skin

In this section, we will discuss some details about the optical properties of human skin. These optical properties are useful to understand the appearance of human skin in different spectrum such as VIS and NIR spectrum. We will also introduce some related research on analysis of human skin.

3.1 Physiology of Human Skin

Human skin is a heterogeneous medium made up of several distinct layers with different optical properties [32]. Figure 3.1 shows the cross sectional diagram of human skin. At macro level, human skin can be divided into three primary layers, which are epidermis, dermis and hypodermis. Function and characteristics of each layer are briefly explained in this section.

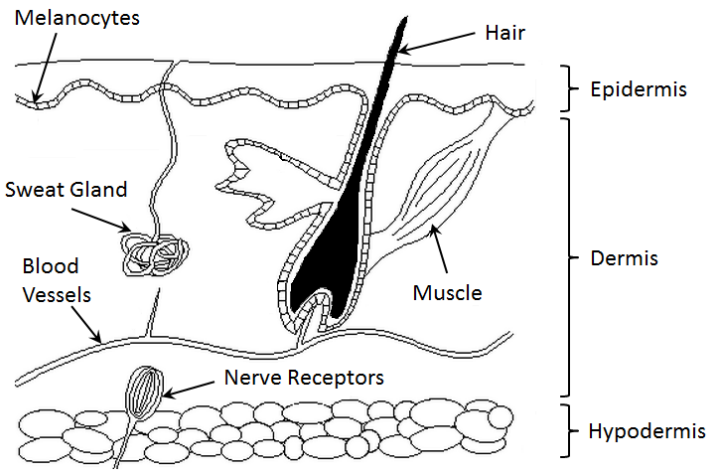

Figure 3.1: Cross-sectional diagram of human skin.

Epidermis

The epidermis layer is the outermost layer of skin. Its thickness varied depending on the location on the body. It is thinnest on the eyelids at 0.05mm and thickest on the soles at about 1.5mm. Generally, it is about 0.1mm in average on face region. The epidermis is further divided into several layers in micro level and is largely composed of connective tissue. However, in this thesis, we will only discuss on macro level.

There are no veins and capillaries in epidermis layer. It is mostly composed of melanocytes, which is the cells that produce melanin. The absorption property of epidermis layer is mostly influenced by melanin's absorption level. The melanin absorption level depends on concentration of melanosome, an organelle containing melanin which absorbs light, in the epidermis. The higher the volume fraction of the epidermis occupied by melanosomes, the darker the skin color. Within the epidermal layer there is very little scattering, with the small amount that occurs being forward directed. All light not absorbed by melanin can be considered passing into the dermis.

Dermis

The dermis layer is located below epidermis layer. This layer is much thicker than the epidermis. Similar to epidermis, its thickness varied depending on the location on the body. For instance, it is about 0.3mm on the eyelids and 1.2mm on the nose.

Dermis is primarily composed of dense and irregular connective tissue with nerves and blood vessels. Specifically, dermis layer is further divided into two layers, which are papillary layer and reticular layer. The papillary layer contains a thin arrangement of collagen fibers which behave as backscattering layer. Scattering is greater in red spectrum and becomes greater in infrared spectrum. The reticular region is composed of thick collagen fibers that are arranged parallel to the surface of the skin. Blood vessels which contain hemoglobin run through both layers of dermis. Hemoglobin is the pigment that gives red color to blood. Compared to the epidermis, there are lower concentration of melanin and higher concentration of hemoglobin in the dermis.

Hypodermis

The hypodermis layer is the layer lies under dermis layer. It is usually not considered as part of the skin. It is a layer of fat and connective tissue that houses a large amount of blood vessels and nerves. It can be up to 3cm thick in the abdomen and absent in the eye lids.

The absorption of hypodermis under VIS spectrum is negligible because VIS photons do not penetrate to this layer. Although NIR light penetrate deeper, scattering of near infrared light is relatively negligible compare to dermis layer.

3.2 Pigments of Human Skin

As introduced in the previous section, human skin can be modeled as a multi-layers structure. Each layer includes various types of light absorbing chemical compounds called pigments. Each layer displays different reflectance properties under different spectral because of their varied ratios of pigments in it. Among these pigments, melanin and hemoglobin are especially important for understanding the appearance of human skin since they are predominantly found in epidermis and dermis layers.

Melanin

Melanin is the dominant pigments of the epidermis. Melanin is produced in melanosomes, and diffused into the epidermis layer. Melanin is divided into two types, known as eumelanin and pheomelanin. Eumelanin is a brown-black pigment that usually exists in dark hair. Pheomelanin is a red-brown pigment that is observed in red hair. Usually, healthy individuals have varying degrees of eumelanin in their skin, while pheomelanin only presents in individuals who has certain genetic trait. Figure 3.2 shows extinction coefficients of eumelanin and pheomelanin [33]. Absorption of pigment is proportional to its extinction coefficient. We can see that absorption of both types of melanin are approximately the same. Therefore, in this thesis, we do not intent to separate melanin in to eumelanin and pheomelanin.

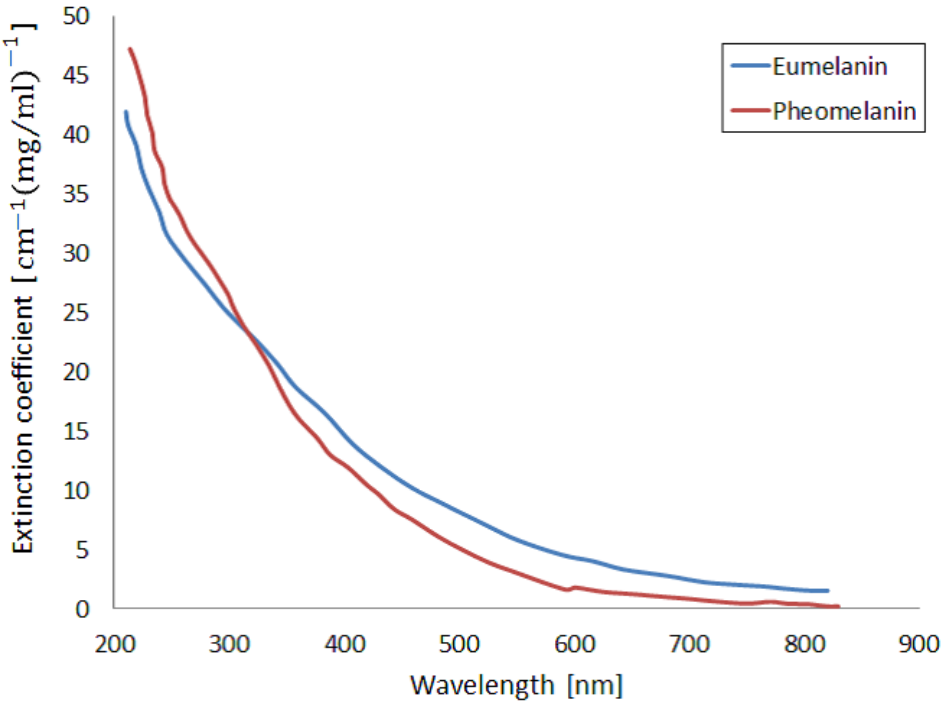

Figure 3.2: Extinction coefficients of eumelanin and pheomelanin.

Melanin is primary determinant of human skin color. It is produced by melanocytes which

found in the epidermis layer. The physiological function of melanin is to protect the human by absorbing and scattering ultraviolet light. Melanocytes start to produce melanin when exposed to ultraviolet radiation of sunlight. Melanin produced is diffused into epidermal layer and eventually moves up towards the surface of skin. This reaction is the phenomenon that makes our skin appears tanned.

The color of skin depends on the volume ratio of melanin in skin. The higher the ratio of melanin, the darker the skin color. In the light colored skin of Caucasians, the ratio is only between 1$\sim$3%. In the skins of well-tanned Caucasians or Asians, the percentage increases to 11$\sim$16%. In dark colored skin of Africans, it can goes up to 43%.

In most individuals, absorption of melanin pigment dominates the absorption properties of the epidermis. In general, the reflectance spectrum of melanin in the visible range is monotonically increasing with wavelength with maximum absorption occurring in the UV range.

Skin features, such as freckles and moles are formed where there is a localized concentration of melanin in the skin. Acne scars which are dark-brown in color are caused by melanin deposit.

Hemoglobin

Hemoglobin is a red colored pigment which can be found in erythrocytes in blood vessels. Hemoglobin binds oxygen effectively and carries oxygen to every parts of body site through vessels and capillaries. When hemoglobin contains oxygen, it is called oxy-hemoglobin. Otherwise, it is called deoxy-hemoglobin.

Blood vessels is composed of arteries that carry bloods from heart to various parts of body and veins that carry deoxygenated blood back to heart. In general, 90–95% of the hemoglobin in arteries and more than 47% of the hemoglobin in veins is oxy-hemoglobin. Most of the hemoglobin in blood vessels are oxy-hemoglobin, therefore, we refer ”oxy-hemoglobin” as ”hemoglobin”. Figure 3.3 shows molar extinction coefficient of oxy-hemoglobin and deoxy-hemoglobin, respectively [34]. Both types of hemoglobin are similar, which in general, have higher absorption in UV range and lower absorption in NIR range. Oxy-hemoglobin exhibit two significant peaks around 500–600nm while deoxy-hemoglobin has only one peak at that range.

Angelopoulou [12] points out that hemoglobin have unique absorption properties around 500–600nm which appear in human skin reflectance properties. Skin reflectance estimated by Angelopoulou versus absorption of hemoglobin measured by Zijlstra *et al.* [35] is plotted in Figure 3.4. Absorption of hemoglobin shows local maximas approximately at 540nm and 575nm. This is corresponding to local minimas of skin reflectance as shown in Figure 3.4. Similarly, local minima of hemoglobin absorption is corresponding to local maxima of skin reflectance at 560nm. Therefore, it is likely that hemoglobin is also a dominant pigment which affects skin appearance. Acne scars which show red or pink in color are caused by hemoglobin deposit.

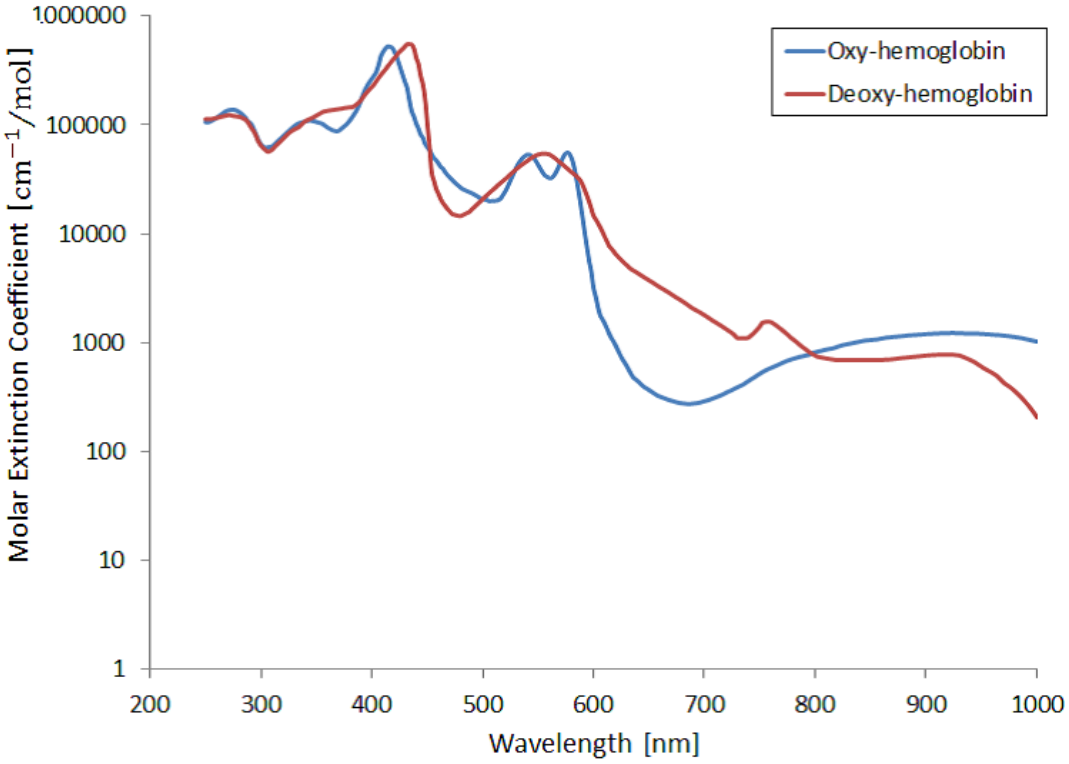

Figure 3.3: Molar extinction coefficients of oxy-hemoglobin and deoxy-hemoglobin.

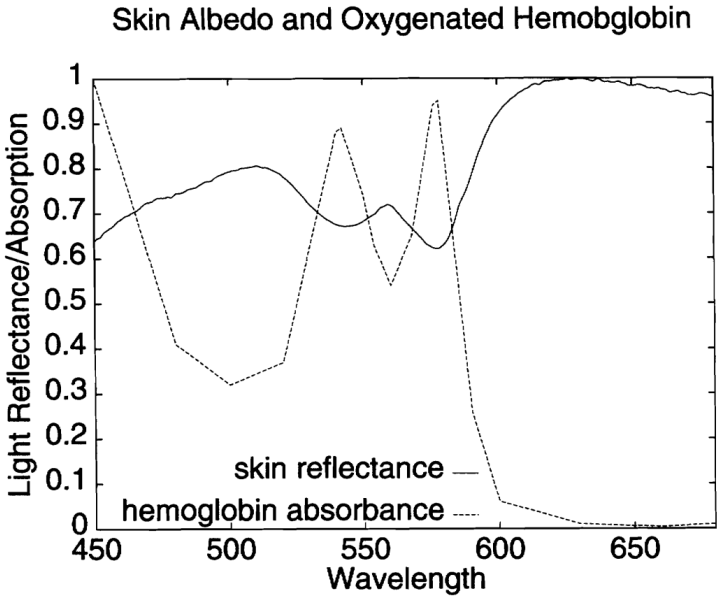

Figure 3.4: Skin reflectance versus hemoglobin absorption [12].

3.3 Analysis of Human Skin Color

The color of human is mainly determined by the absorption and scattering characteristics of skin under VIS spectrum. Nakai *et al.* reported in their work [36] that the spectral reflectance of various skin colors can be derived from absorption characteristic of three pigments found in skin layers which are melanin, hemoglobin, and carotene. They employed multiple regression analysis (MRA) to the spectral reflectance of various skin colors and success in representing the spectral reflectance with only three variables mentioned above. However, since the carotene least common pigment and its concentration is much less compare to melanin and hemoglobin, it is possible to represent skin spectral reflectance using only the absorption properties of melanin and hemoglobin.

Tsumura *et al.* proposed a method to extract melanin and hemoglobin information from a skin color image. Figure 3.5 shows the flow diagram of Tsumura's method of extracting both pigments from a color image and reuse of the pigment information to synthesize various skin colors. The original face image is first separated into the images of body reflection and surface reflection. The surface reflection is specular reflection caused by directional light while the body reflection is the diffuse reflection caused by scattering light in human skin structure. The image of body reflection is decomposed into melanin and hemoglobin component using ICA. In this stage, shading is removed using their proposed shading removal technique to avoid inaccurate estimation of pigment densities. The pigment densities are then changed spatially based on physiological knowledge of skin colors in synthesis phase. Lastly, the synthesized body reflection image is combined with surface reflectance image for final synthesized image.

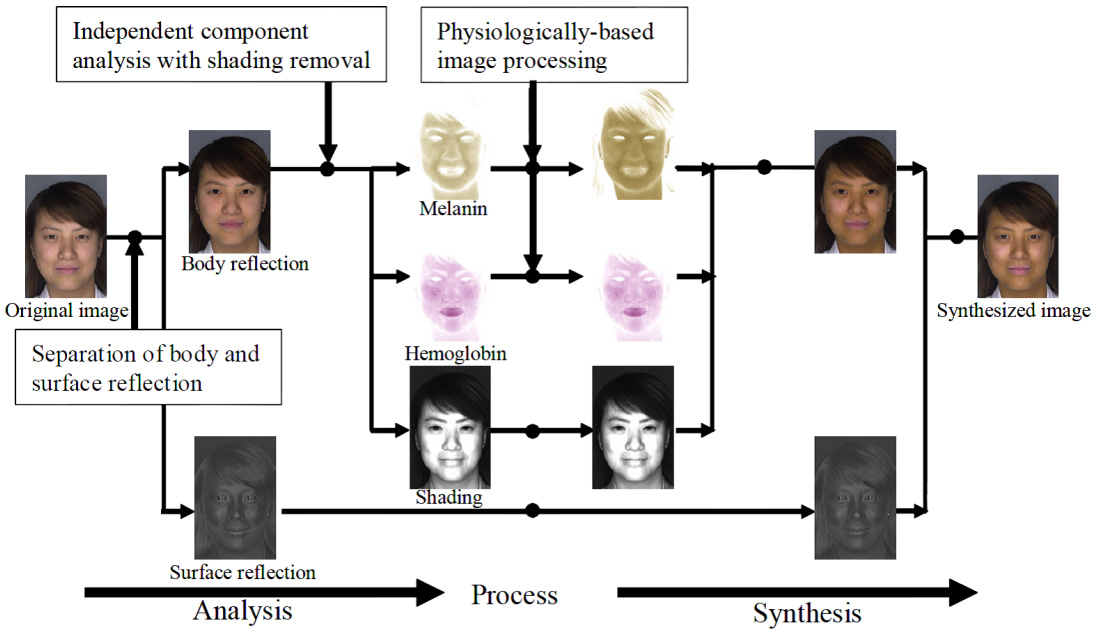

Figure 3.5: Flow diagram for skin color analysis/synthesis proposed by Tsumura *et al.* [13].

In their work, few assumptions are assumed as following:

(1) spatial variation of color in the skin is caused by two pigments, melanin and hemoglobin,

(2) their quantities are mutually independent spatially,

(3) the linearity holds among the quantities and observed color signals in the optical density domain.

By these assumptions, skin color under homogeneous shading area are distributed approximately on a plane spanned by two vectors $\boldsymbol{c}(1)$ and $\boldsymbol{c}(2)$ in the logarithmic RGB space as illustrated in Figure 3.6. The two vectors $\boldsymbol{c}(1)$ and $\boldsymbol{c}(2)$ represent the absorption properties of melanin and hemoglobin, respectively. Skin color $\boldsymbol{c}_{i,j}$ at position (i, j) can be represented as follow

$$\boldsymbol{c}_{i,j} = q_{i,j}(1)\boldsymbol{c}(1) + q_{i,j}(1)\boldsymbol{c}(1) + \boldsymbol{b}, \tag{3.1}$$

where $\boldsymbol{c}(1)$ and $\boldsymbol{c}(2)$ are color vectors of melanin and hemoglobin (or hemoglobin and melanin), $q_{i,j}(1)$ and $q_{i,j}(2)$ are quantities of pigments at position (i, j), respectively, and $\boldsymbol{b}$ is a bias vector caused by light source.

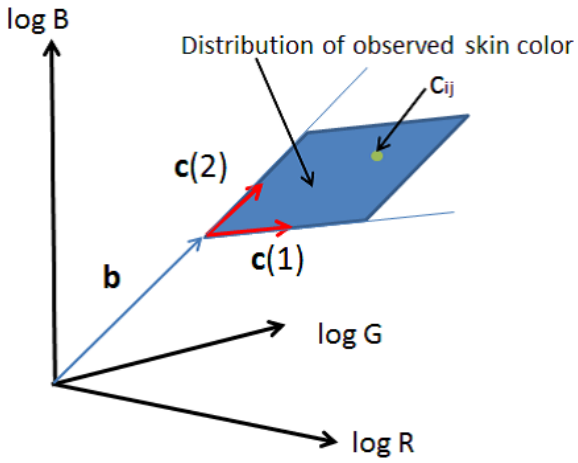

Figure 3.6: Skin colors distribute on a plane spanned by two vectors corresponding to melanin and hemoglobin factors.

Tsumura *et al.* also have confirmed the physiological validity of the proposed method by some practical experiments. Their extraction results are shown in Figure 3.7. The arm in (a) is irradiated by UV-B for simulated tanning effect while the arm in (b) is partially applied with methyl nicotinate to improve hemoglobin concentration. The decomposition results of ICA in both cases show that estimated pigment distribution maps agree well with the physiological facts. In (a), first independent component (second row) shows square patterns cause by tanning, which indicates biological fact of melanin. On the other hand, second independent component

(third row) in (b) shows the round pattern where methyl nicotinate applied. This again agrees with the biological response of hemoglobin.

Their analysis was done on only RGB spectral using color image and no further discussion was made for NIR spectrum. Melanin and hemoglobin also influence the skin appearance on NIR spectra [37]. Our method attempts to extract those pigment information from multispectral NIR input and utilized for synthesis purpose.

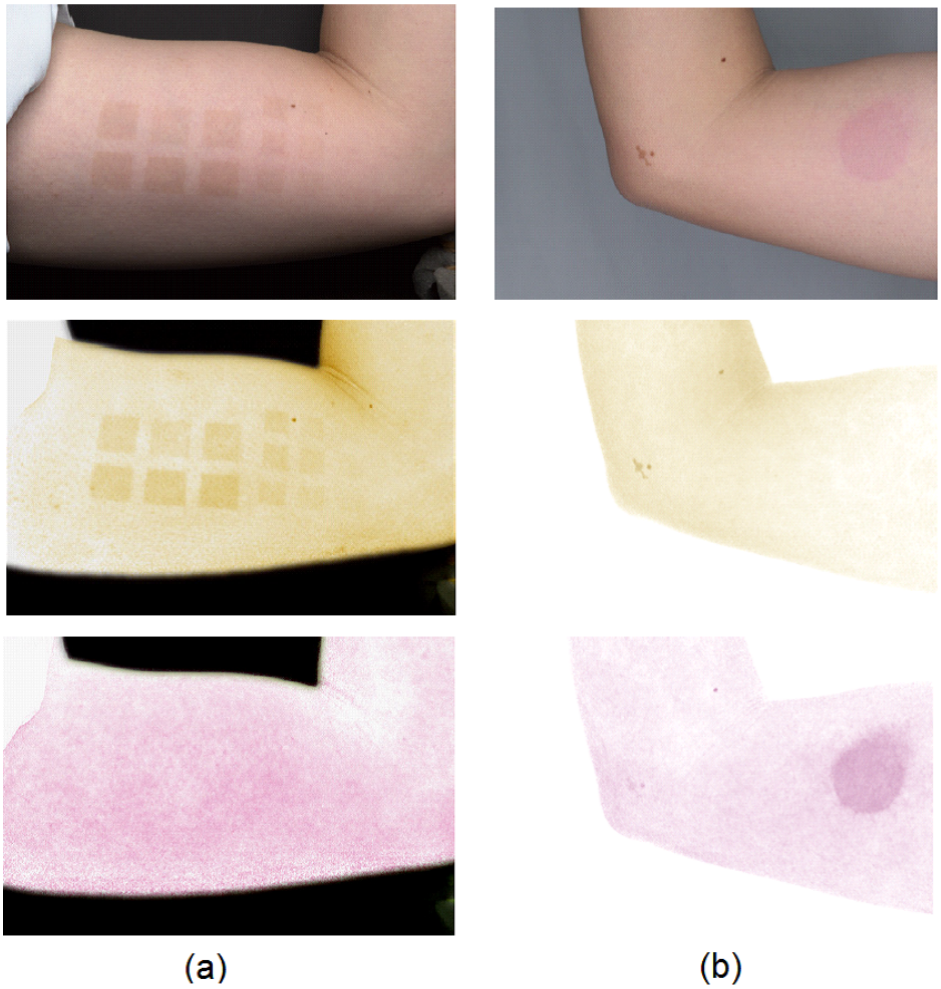

Figure 3.7: Melanin/hemoglobin extraction from Tsumura's method. First row shows original color image, second row shows result for melanin component and third row show result for hemoglobin component. (a) UV-B irradiation, (b) Methyl nicotinate application [13].

Chapter 4

Proposed method

4.1 Skin Pigment Model

As introduced in the previous chapter, the appearance of human skin changes with respect to subsurface scattering of dermal and epidermal layers [11, 12, 38]. To reconstruct the appearance of facial image taken under different wavelength of light, we need to consider subsurface scattering of dermal and epidermal layers in our imaging model.

Figure 4.1 shows a light transport model of epidermal and dermal layers, where subsurface scattering of both layers are well-considered. Melanin and hemoglobin pigments are predominantly contained in these layers. In epidermal layer, part of incoming light is absorbed and scattered by melanin. Light that is not absorbed by melanin pass into the dermal layer and is later absorbed or scattered by hemoglobin before comes out as observed outgoing light. The subsurface reflectance of these layers are well modeled by modified Lambert-Beer law [13] as,

$$L(\lambda_n) = \exp\{-q_m r_m(\lambda_n) - q_h r_h(\lambda_n)\} E(\lambda_n), \tag{4.1}$$

where λ_n is the n-th wavelength of incoming/outgoing light, $E(\lambda)$ and $L(\lambda)$ are the spectral distributions of incoming irradiance and outgoing radiance, respectively, and q_m, q_h, $r_m(\lambda)$, $r_h(\lambda)$ are the pigment densities and absorbance coefficients of melanin and hemoglobin, respectively.

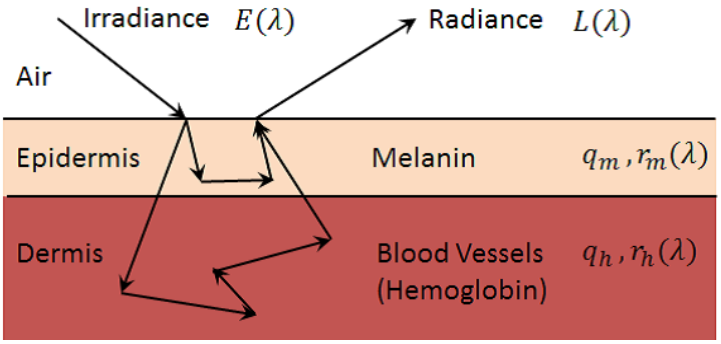

Figure 4.1: Light transport model of epidermal and dermal layers.

Taking the logarithm of Equation (4.1), we obtain the following additive form,

$$c(\lambda_n) = q_m r_m(\lambda_n) + q_h r_h(\lambda_n) + b(\lambda_n), \tag{4.2}$$

where $c(\lambda) = -\log(L(\lambda))$, and $b(\lambda) = -\log(E(\lambda))$.

Equation (4.2) shows the linear relationship of irradiance on logarithmic domain and melanin and hemoglobin pigments. If we observe a certain surface of a skin with N spectral wavelength, irradiance of each pixel on the surface $\boldsymbol{c}_{i,j}$ is an observation in the form of a vector of N dimensions. According to Equation (4.2), the N-dimensional multispectral observation of logarithmic radiance $[c(\lambda_1) - b(\lambda_1), c(\lambda_2) - b(\lambda_2), ..., c(\lambda_N) - b(\lambda_N)]$ lies in two dimensional subspace spanned by $[r_m(\lambda_1), r_m(\lambda_2), ..., r_m(\lambda_N)]$ and $[r_h(\lambda_1), r_h(\lambda_2), ..., r_h(\lambda_N)]$. This distribution is illustrated as in Figure 4.2. Note that $\boldsymbol{r}_m = [r_m(\lambda_1), r_m(\lambda_2), ..., r_m(\lambda_N)]^t$ and $\boldsymbol{r}_h = [r_h(\lambda_1), r_h(\lambda_2), ..., r_h(\lambda_N)]^t$ are the basis that represents absorption properties of melanin and hemoglobin at corresponding wavelength.

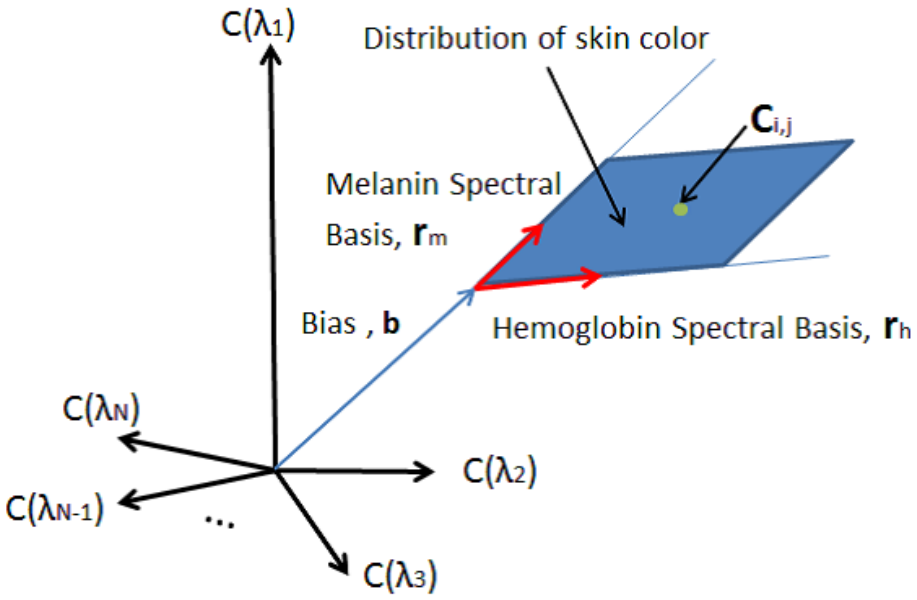

Figure 4.2: Distribution of skin irradiance in the N-dimensional multispectral space.

4.2 Overview of Proposed Method

Our method is based on the skin pigment model introduced in previous section, which is applicable in both VIS and NIR spectra. The conceptual diagram of the proposed method is shown in Figure 4.3.

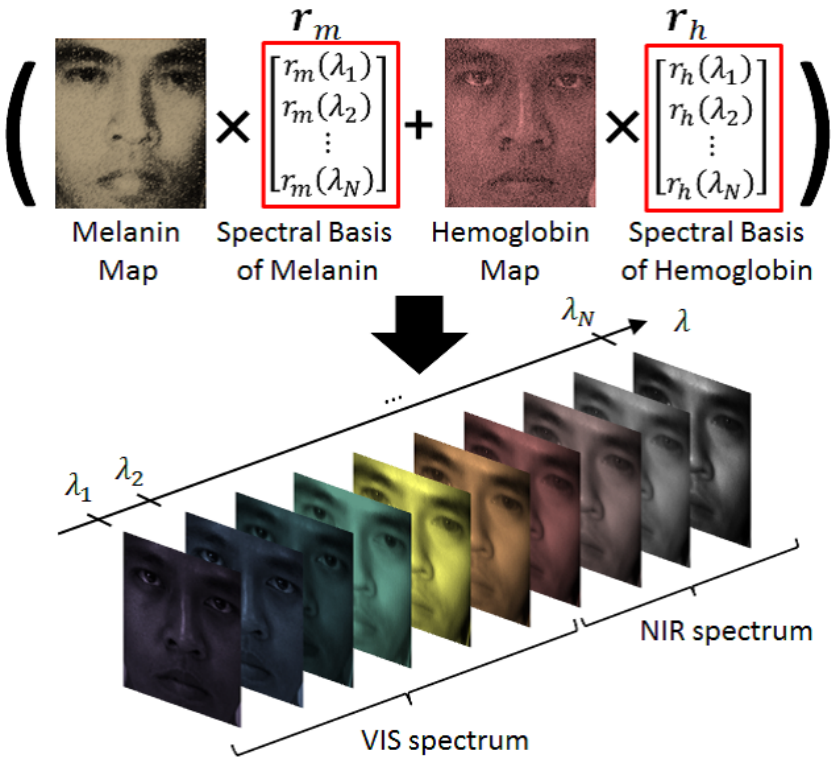

Figure 4.3: Conceptual diagram of the proposed method.

We use multispectral images which are taken under N different wavelengths that cover VIS and NIR spectrum. We denote logarithmic of N pixel values obtained along wavelength axis at position (i, j) as a column vector, $\boldsymbol{c} = [c(\lambda_1), c(\lambda_2), ..., c(\lambda_N)]^t$. Expanding Equation (4.2) to N spectral images, we obtain follows,

$$\begin{bmatrix} c(\lambda_1) \\ c(\lambda_2) \\ \vdots \\ c(\lambda_N) \end{bmatrix} = \begin{bmatrix} r_m(\lambda_1) & r_h(\lambda_1) \\ r_m(\lambda_2) & r_h(\lambda_1) \\ \vdots & \vdots \\ r_m(\lambda_N) & r_h(\lambda_1) \end{bmatrix} \begin{bmatrix} q_m \\ q_h \end{bmatrix} + \begin{bmatrix} b(\lambda_1) \\ b(\lambda_2) \\ \vdots \\ b(\lambda_N) \end{bmatrix}. \tag{4.3}$$

Then, we rewrite Equation (4.3) by vector and matrix formulation as follows:

$$\boldsymbol{c} = \boldsymbol{R}\boldsymbol{q} + \boldsymbol{b}, \tag{4.4}$$

where $\boldsymbol{R} = [\boldsymbol{r}_m, \boldsymbol{r}_h]$, $\boldsymbol{r}_m = [r_m(\lambda_1), r_m(\lambda_2), ..., r_m(\lambda_N)]^t$, $\boldsymbol{r}_h = [r_h(\lambda_1), r_h(\lambda_2), ..., r_h(\lambda_N)]^t$, $\boldsymbol{q} = [q_m, q_h]^t$, and $\boldsymbol{b} = [b(\lambda_1), b(\lambda_2), ..., b(\lambda_N)]^t$ is a constant bias vector caused by irradiance.

Since $\boldsymbol{r}_m$ and $\boldsymbol{r}_h$ are the bases that describe the absorption coefficients of melanin and hemoglobin at specific wavelength, we call these bases as spectral basis. Spectral basis is invariant to the subjects in facial images. Our method learns spectral basis of melanin and hemoglobin from a multispectral image dataset. Then, given a multiband NIR image, we can estimate the coefficients, which are pixel-wise densities of melanin and hemoglobin. Using the pixel-wise densities of pigments estimated and spectral basis, we can synthesize facial images in VIS wavelength.

Figure 4.4 shows the schematic flow of our method composed of 3 main steps; these are spectral basis learning, pigment densities estimation, and VIS images synthesis. Further explanation of each step is described in next section.

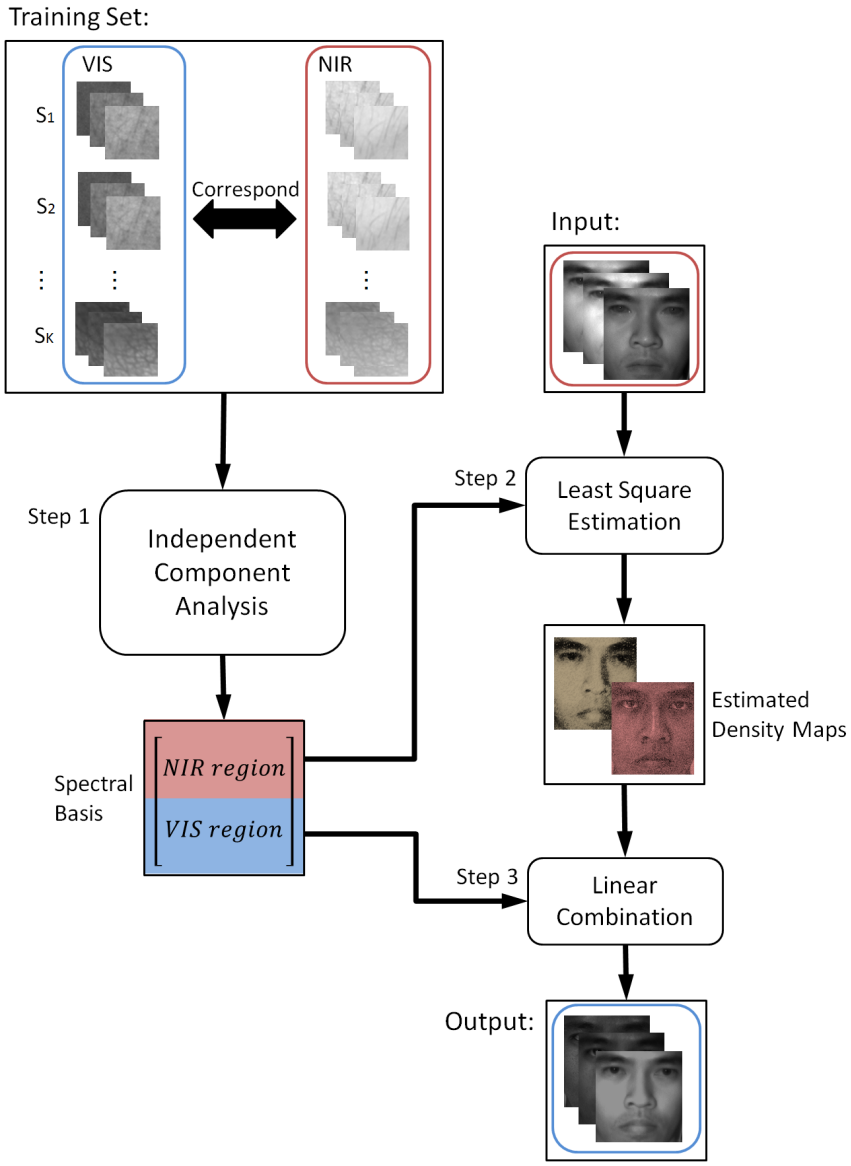

Figure 4.4: Schematic flow of the proposed method.

4.3 Learning Spectral Basis

A dataset composed of N spectral images of various subjects are first acquired from VIS and NIR spectra. Skin patches with size of $W \times H$ are clipped from K subjects to form the training set. For each spectral, patch is expressed as a $P(= WH)$ dimensional column vector. Then patches from all subjects are concatenated to form a N–by–PK data matrix $\boldsymbol{C}$. Figure 4.5 shows the flow chart of the decomposition of this data matrix into two matrices which are corresponding to pigment densities and absorbance coefficients.

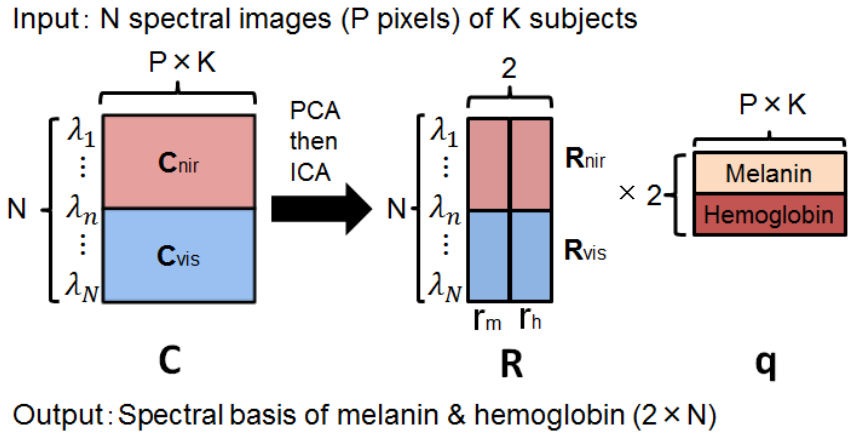

Figure 4.5: Flow chart of data matrix being decomposed into melanin and hemoglobin factors.

To extract two-dimensional subspace of the dominant pigments, principal component analysis (PCA) is firstly applied [13]. Figure 4.6 shows the cumulative contribution ratio of 6 principal components obtained from the 6 spectral data used in the experiment. This figure shows that two principal components are sufficient to describe the values of 6 spectra with accuracy as high as 96.8%.

Then, the input data is projected onto the two-dimensional subspace spanned by eigenvectors corresponding to first and second principal components.

We then perform ICA [14] on projected data to decompose the data matrix into 2 matrices, which are mixing matrix $\boldsymbol{R}$ and densities matrix $\boldsymbol{q}$. Note that mixing matrix $\boldsymbol{R}$ is composed of spectral basis of melanin $\boldsymbol{r}_m$ and hemoglobin $\boldsymbol{r}_h$, that describe absorbance coefficient of both pigments at specific wavelength. Many approaches have been proposed from ICA, in our case, we perform ICA by mutual information minimization approach [39].

Each spectral basis vector is divided into two subvectors based on NIR and VIS spectrum region. Subvector on NIR region $\boldsymbol{R}_{nir}$ is used for estimating density map of melanin and hemoglobin. On the other hand, subvector on VIS region $\boldsymbol{R}_{vis}$ is used in synthesis step. Similarly, the bias vector obtained from the training set is divided into subvectors and used for following steps.

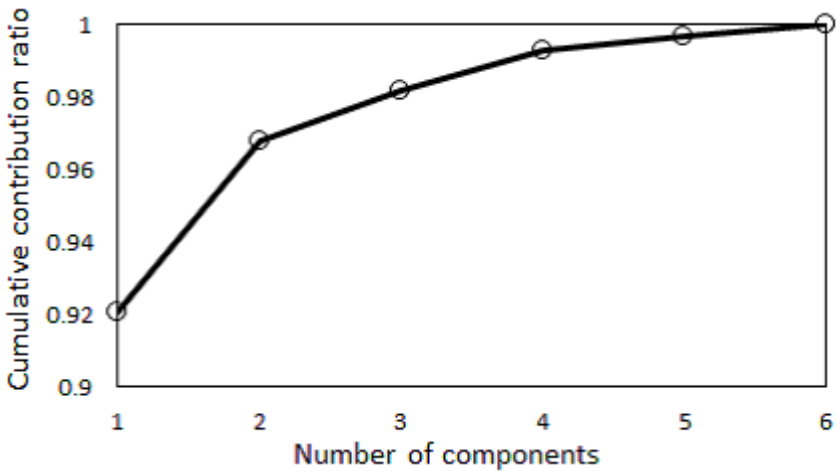

Figure 4.6: Relationship between the number of components and the cumulative contribution ratio in skin image set of 6 spectra.

4.4 Estimation of Density Maps

Given at least two NIR images taken under different wavelengths, density map of melanin and hemoglobin densities can be estimated by minimizing the reconstruction error of NIR images. For the error minimization task, we use least square estimation as follows:

$$\widehat{\boldsymbol{q}}(i,j) = \underset{\boldsymbol{q}(i,j)}{\operatorname{argmin}} \parallel \boldsymbol{c}_{nir}(i,j) - \boldsymbol{R}_{nir}\boldsymbol{q}(i,j) - \boldsymbol{b}_{nir} \parallel^2, \tag{4.5}$$

where $\widehat{\boldsymbol{q}}(i,j)$ is the estimated pigment densities at position (i,j) on image, $\boldsymbol{c}_{nir}$ is logarithmic values of input NIR images, and $\boldsymbol{b}_{nir}$ is the subvector of bias vector $\boldsymbol{b}$ in NIR region.

4.5 Synthesizing VIS Images

For synthesis step, we use the estimated $\widehat{\boldsymbol{q}}$ and $\boldsymbol{R}_{vis}$ in Section 3.1 to synthesize images in VIS region. The synthesis process is indicated as

$$\boldsymbol{c}_{vis}(i,j) = \boldsymbol{R}_{vis}\widehat{\boldsymbol{q}}(i,j) + \boldsymbol{b}_{vis}, \tag{4.6}$$

where $\boldsymbol{c}_{vis}$ is logarithm values of output VIS images, and $\boldsymbol{b}_{vis}$ is the subvector of bias vector in VIS region. We then find the inverse log of $\boldsymbol{c}_{vis}$ and reshape it back to image size.

Since we estimate and synthesis in a pixel-wise manner using common spectral basis of melanin and hemoglobin over a face, registration of face patch is not needed. Therefore, our method can be also applicable to facial images with different expressions and orientations to the training set.

Chapter 5

Experiments

5.1 Image Acquisition and Preprocessing

In this thesis, we set N=6, where we use 3 NIR images to synthesize 3 VIS images (RGB-channels) to confirm the validity of our method. Three NIR images are used instead of two to reduce the density estimation error. Figure 5.1 shows the multispectral imaging system used in our experiments. This multispectral imaging system composed of a camera and a filter wheel with 6 band-pass filters attached to it. The camera used is Chameleon CMLN-13S2M-CS which has spectral sensitivity range from 400nm to 1000nm. Band-pass filters used are of narrowband range with center wavelength at 450nm, 532nm, 610nm, 766nm, 880nm, and 960nm, respectively. We use halogen lamp for our light source. To ensure our multispectral imaging system capture sharp images for both VIS and NIR spectrum, we use varifocal len (13FM08IR) for our experiments.

For our experiment uses, a database composed of 8 subjects are acquired at 6 wavelengths mentioned above. Face images of the same subject taken under different wavelength are manually aligned. Normalization of face images between different subjects is not required. Face region of images are clipped into the size of 200×200. Training set is composed of skin patches clipped from various face region of frontal facial images with neutral expression. For evaluation, facial images of the same subjects include different orientations and expressions to training set are acquired, and we learn spectral bases from different subjects to input NIR images. We used MATLAB R2011b as our software for image processing throughout our whole experiments.

Figure 5.2 shows an example of spectral images of a subject taken under 6 wavelengths with our multispectral imaging system after preprocessing.

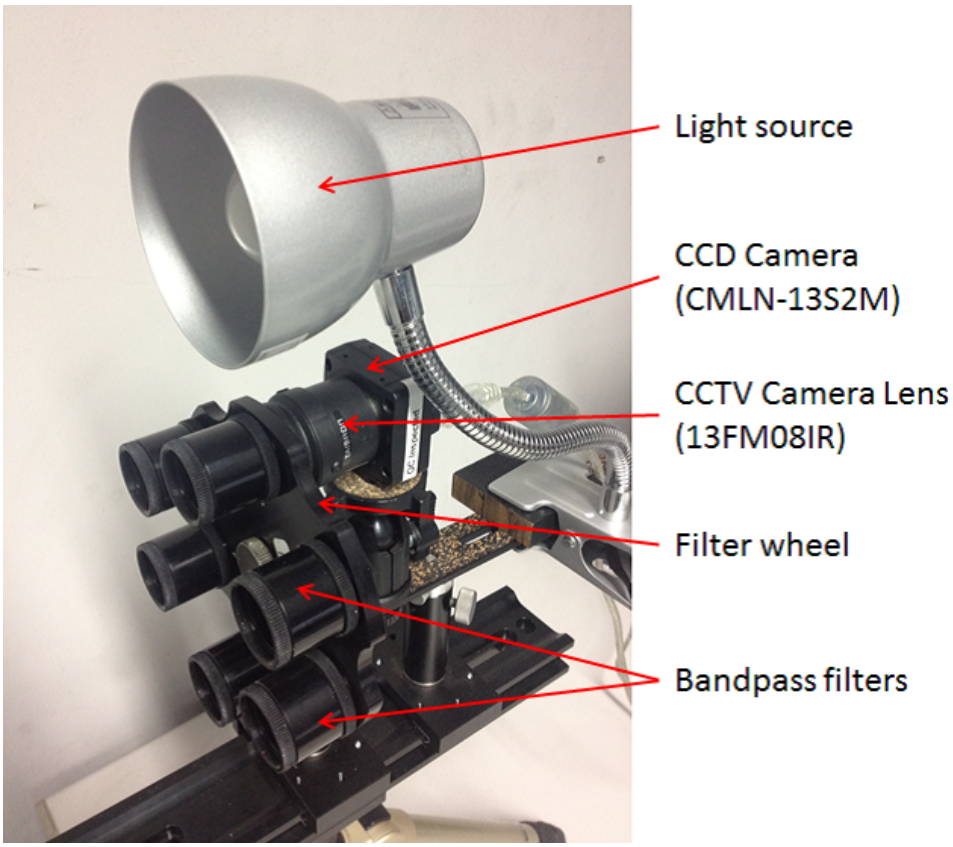

Figure 5.1: Multispectral imaging system in our experiments.

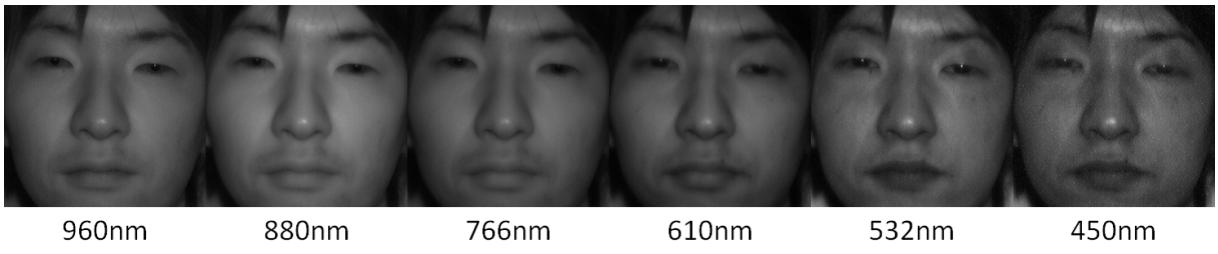

Figure 5.2: Example of 6 spectral images of a subject after preprocessing.

5.2 Evaluation of Spectral Basis Learning

Spectral basis describe absorption coefficients of melanin and hemoglobin at specific wavelength which are invariant to subject in facial images. We learn these spectral bases from a training set which include facial image samples of different subjects taken under different wavelength. When there is insufficient number of samples in training set, spectral basis learned is unreliable due to overfitting problem which is commonly seen in machine learning approach. In this section, two experiments is designed to evaluate the efficiency and validity of spectral basis learning, respectively.

First experiment is to evaluate the efficiency of spectral basis learning. Generally, overfitting problem can be avoid by increasing the size of training set. However, increasing the size of training set will costs more computational times which also means the decreasing of learning efficiency. In this experiment, we evaluate the efficiency of spectral basis learning by determining the minimum number of subjects need to be involved in training set for a successive learning process. We leave out one subject from the database as test subject while other subjects are used for training set. Then, we change the number of subjects to be included in training set and learn a set of spectral basis from it. Spectral basis learned are used to reconstruct VIS images of the test subject, given NIR images as input. Spectral basis is considered reliable when the reconstruction error is small. Figure 5.3 shows the mean absolute errors of synthesized images versus number of subjects to be involved in training set. Errors are evaluated by averaging the absolute difference between synthesized image and ground truth image of all pixels. Errors of the synthesized images are reduced as we increase the number of subjects involved in training set. From the result, we can induce that the use of skin patches clipped from 2–3 subjects are sufficient for our learning process.

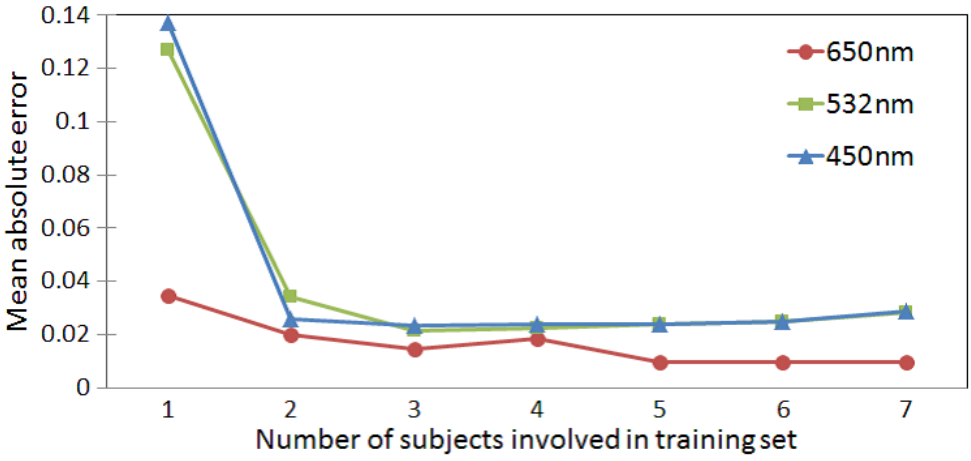

Figure 5.3: Mean absolute error of synthesized images in different number of subjects involved in training set.

In second experiment, we evaluate the validity of our spectral basis learning. We randomly select 3 subjects from our database to form a training set. We repeat this process to cover all combination of selecting 3 subjects from 7 subjects. These training set are used for spectral basis learning. Mean and standard deviation value of each element in spectral basis learned for each wavelength are determined. The results are shown in Figure 5.4.

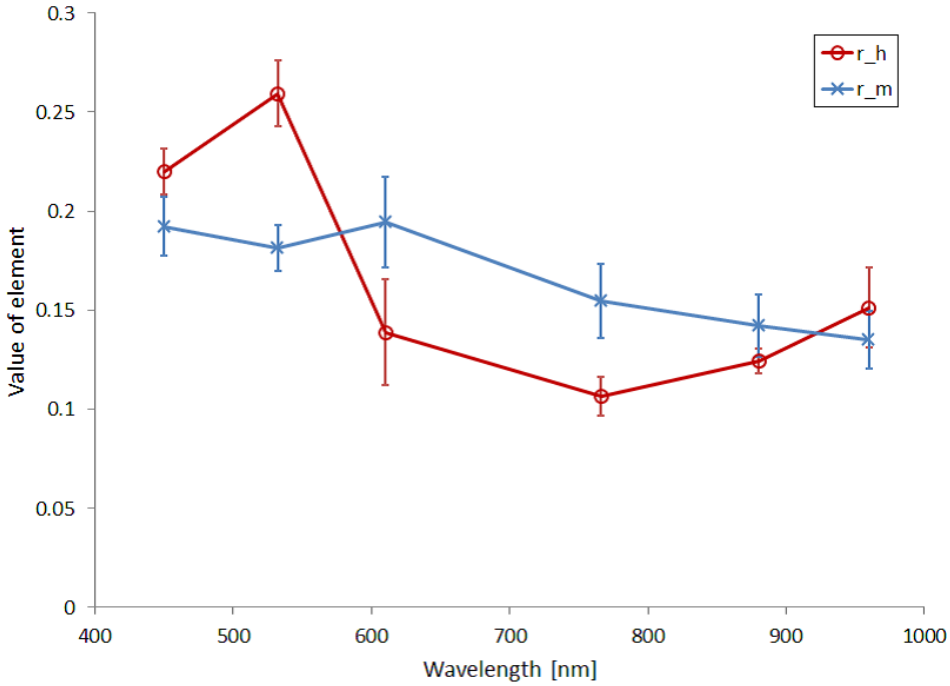

Figure 5.4: Values of elements in spectral basis learned in different training set. Error bars indicates the standard deviation of element value at specific wavelength.

Elements in each spectral basis represent the absorption properties of melanin and hemoglobin at specific wavelength. Comparing our result for hemoglobin spectral base with Figure 3.3 in Section 3.2, we can see that the result match well with the actual extinction coefficients of oxy-hemoglobin measured. Elements in spectral base of hemoglobin learned shows a high values around 500–550nm in VIS spectrum, gradually decreases when moving toward NIR spectrum, and slightly increases again in NIR spectrum.

Melanin is known to have exponential dependence on wavelength with higher absorption at shorter wavelength [40]. The absorption of melanin is gradually decreases when moving towards NIR spectrum. Our result for melanin spectral base exhibit the properties of melanin absorption mentioned above. Although the spectral basis of melanin learned exhibit a relatively higher value at 600nm, the estimation is still acceptable if we consider the standard deviation of the element.

5.3 Pixel-wise Pigment Densities Estimation

Figure 5.5 shows the result of pigment densities estimation based on two learned spectral bases of melanin and hemoglobin. Input NIR multiband images used are shown in Figure 5.6. Lower intensity of the image indicates lower pigment density. Figure 5.5(a) shows lower concentration of pigment at the lip region and higher concentration of pigment at mole spot on the right cheek. This estimation result agrees well with the physiological facts of melanin [11]. Although mole spot in input NIR images shown in Figure 5.6 do not display much differences in intensity, our method still able to estimate pigment densities at relevant spot correctly. Similarly, Figure 5.5(b) shows higher concentration of hemoglobin pigment on lip region. However, because the pigment model holds only for skin region, pigment densities at the region of eyes and eyebrows are not estimated correctly.

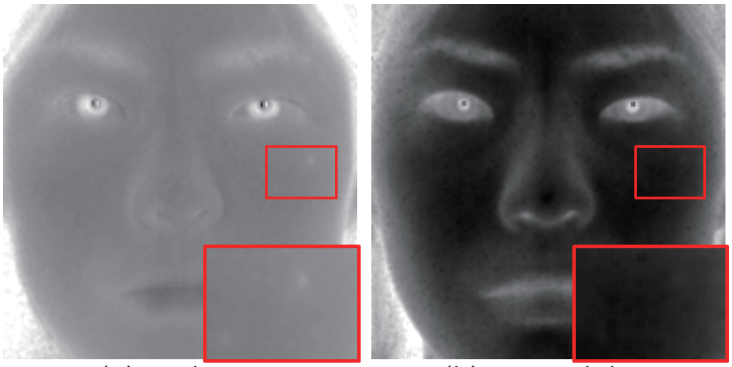

Figure 5.5: Estimated pigment densities corresponding to two independent components.

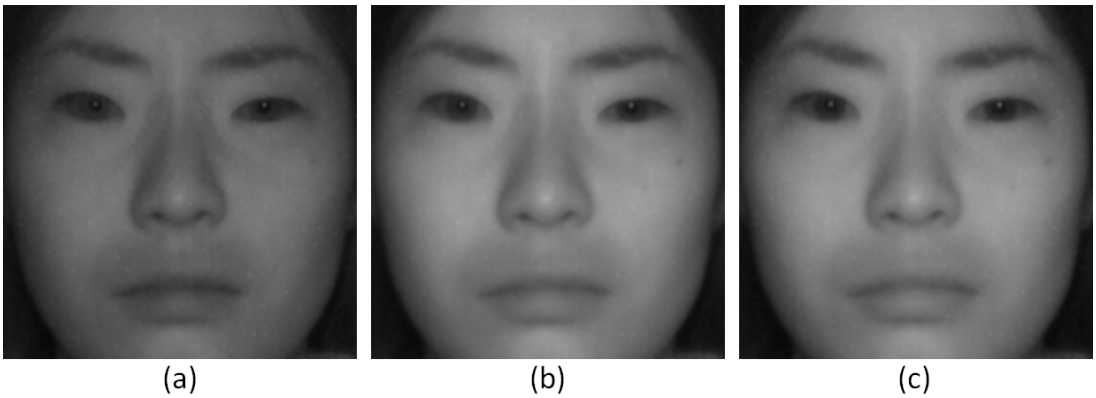

Figure 5.6: Input NIR images for pigment densities estimation. (a) 960nm, (b) 880nm, (c) 766nm.

5.4 VIS Images Synthesis

Figure 5.7 shows synthesized VIS images by using two pigment density maps shown in Figure 5.5. Error of the result is evaluated by calculating the different of pixel intensity between synthesized result and ground truth image and emphasized for visualization.

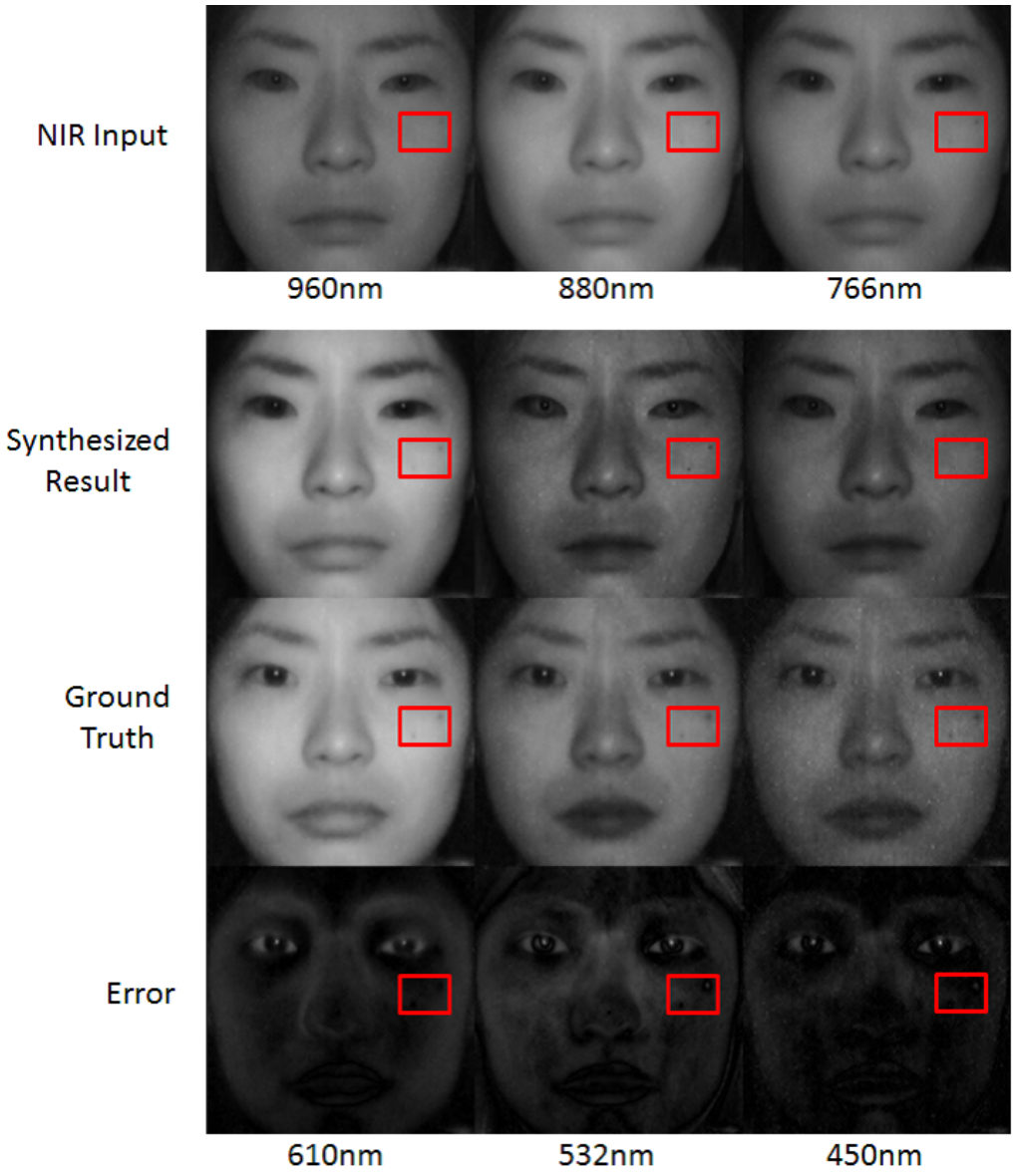

Figure 5.7: Synthesized VIS facial images of frontal view.

We can see that appearance of skin region is visually close to the ground truth and the errors are tolerable. However, the appearance of eyes region is not synthesized to be what expected under VIS spectrum. This occurred because our pigment model holds only for skin region. It can be refined by separating eyes region from facial images and applying different synthesis

algorithm for that region.

Close-ups of the mole spot (red bounding boxes) are shown in Figure 5.8. Synthesized image for wavelength at 610nm and 532nm look fine with details such as mole on the right cheek preserved. However, mole spot in synthesized image for wavelength at 450nm almost disappear. This occurs because of the element of melanin spectral basis learned for wavelength at 450nm is insubstantial to describe absorption of melanin as addressed in Section 5.2.

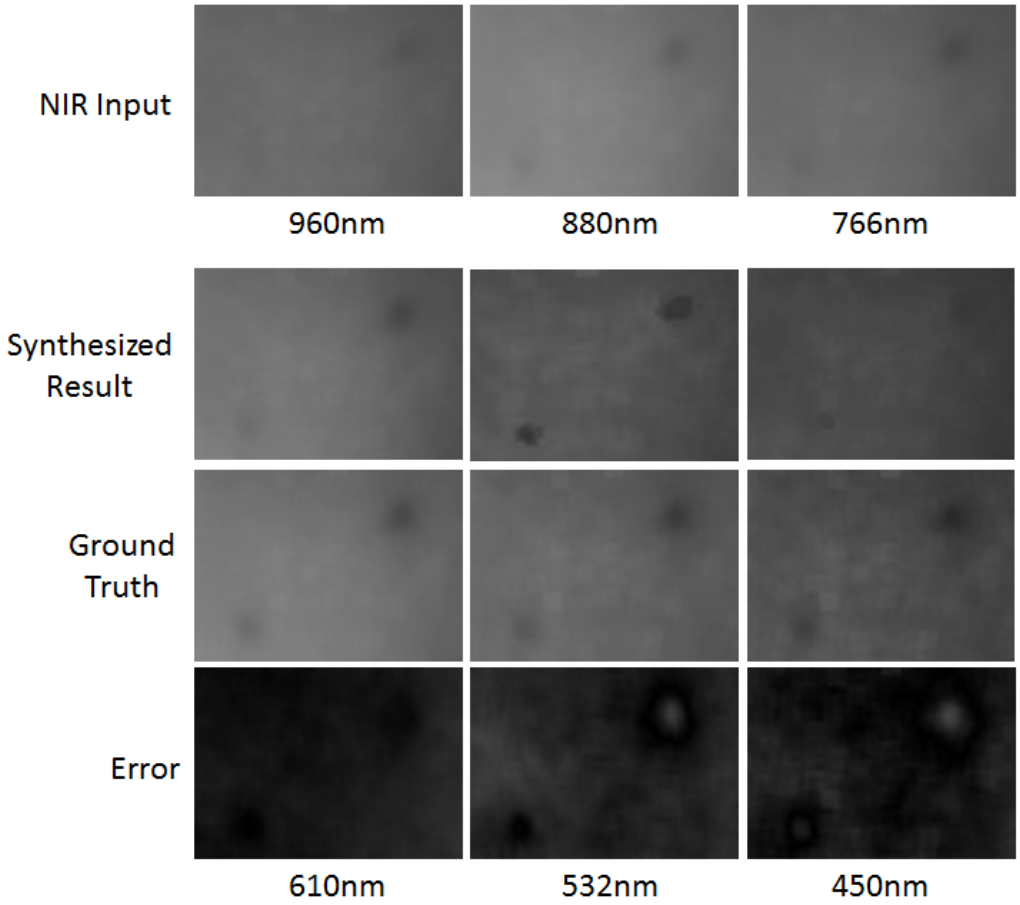

Figure 5.8: Close-ups of red bounding boxes in synthesized VIS images of Figure 5.7.

Figure 5.9 shows comparison result of proposed method with a naive method using multivariate linear regression method (MLR). In linear regression model, the relationship between the VIS and the vector formed by multiband NIR images is assumed linear. We use the same training samples for both proposed method and MLR method. We can see that the error of MLR method is greater than proposed method overall.

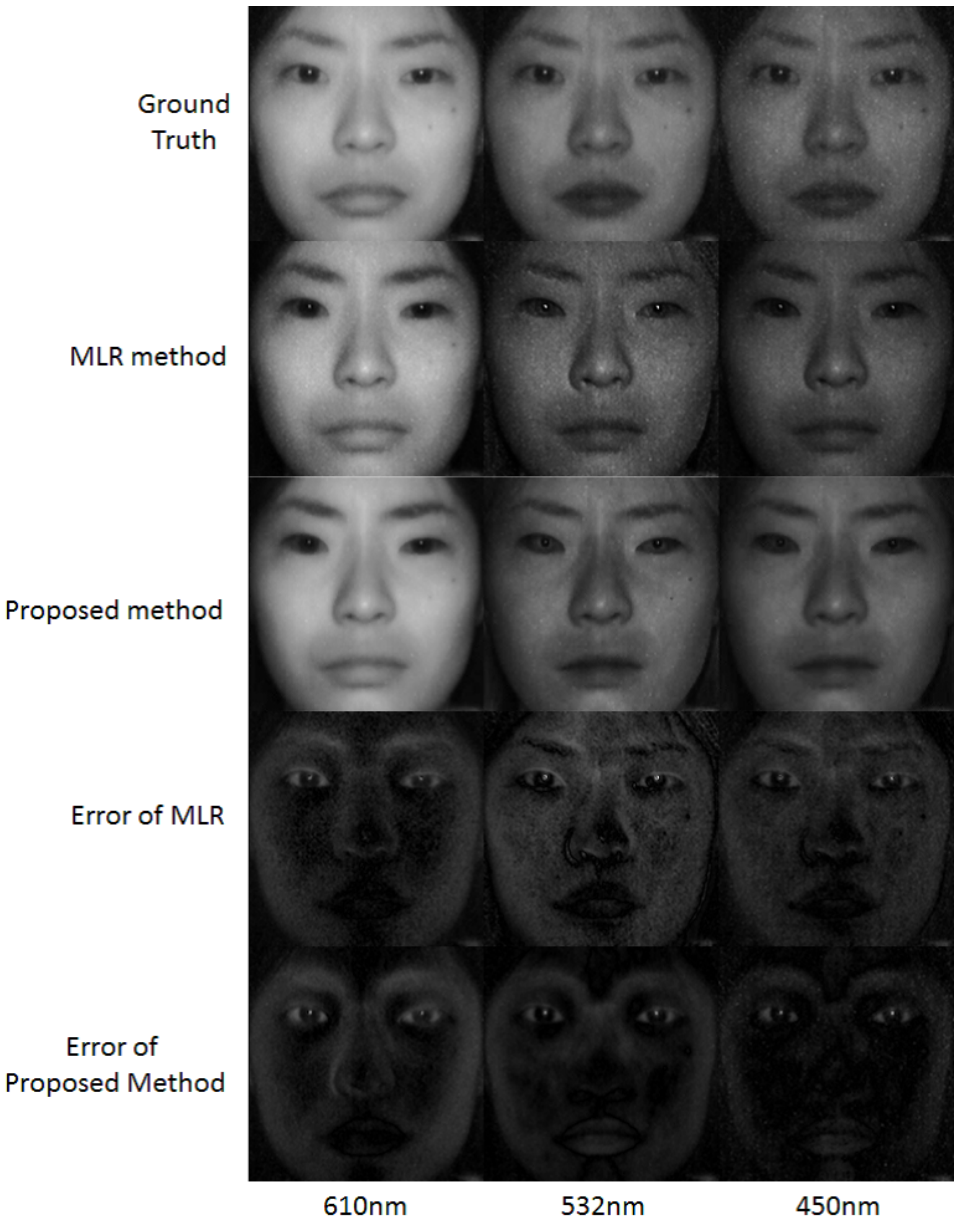

Figure 5.9: Comparison of proposed method with naive method using multivariate linear regression.

5.5 Synthesis on Images with Different Orientation and Expression

We have also tested our synthesis method on NIR input with different orientation and expression. Synthesized results for probe images of side view and with expression are shown in Figure 5.10 and Figure 5.11, respectively. The errors is generally greater compared to synthesis experiments with frontal image with neutral expression. However, the errors are tolerable and the synthesized VIS images are visually well reconstructed. Note that even though our training

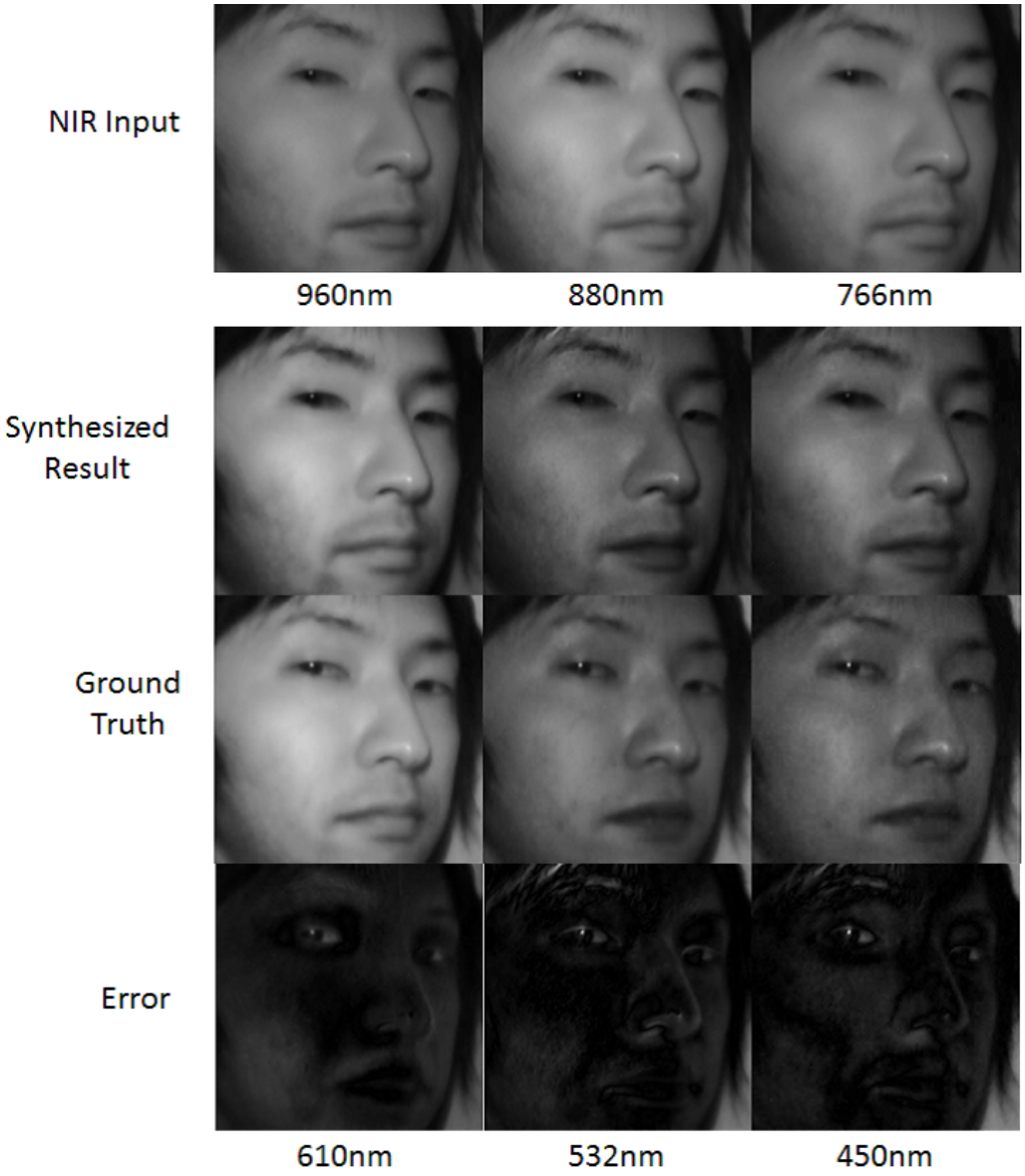

Figure 5.10: Synthesized results using facial images of side view.

set contains only skin patches of frontal view and neutral expression, our method is able to con-

vert probe NIR images of different orientations and expressions to RGB images. This is hard to be achieved by patch-based synthesis method.

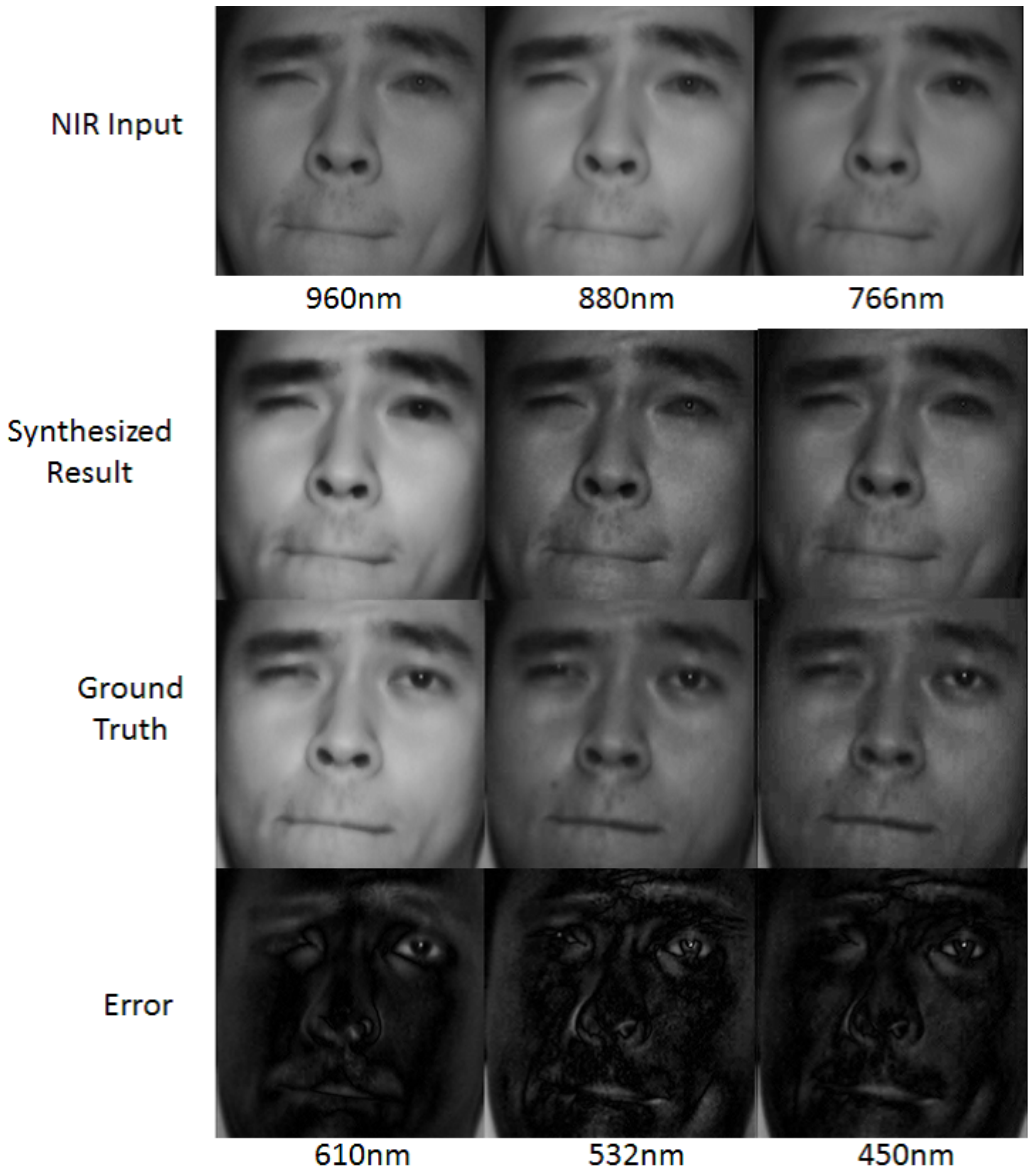

Figure 5.11: Synthesized results using facial images of different expression.

Chapter 6

Conclusion

6.1 Summary

We have presented a method for converting facial images from multi-band NIR images to VIS images based on photometric properties of human skin. Spectral reflectance of human skin is predominated by absorption properties of melanin and hemoglobin.

Our basic idea is to learn spectral basis from a small training set with skin patches. Spectral basis is interpersonally invariant. Appearance of human skin changes with respect to amount of melanin and hemoglobin contained under skin. We utilize spectral basis learned to estimate densities of melanin and hemoglobin in pixel-wise manner. Artificial VIS images are synthesized using linearity of pigment quantities and the observed signal in optical densities domain.

Experimental results showed that our method works well for real facial images. Our training set requires just skin patches from few subjects for learning spectral basis. Spectral basis agree well with the absorption properties of pigments. We succeeded in estimate pixel-wise densities of melanin and hemoglobin, given multiband NIR images as input. Important facial image details, such as mole are well preserved in our synthesized VIS images. Our synthesized VIS images are visually close to the ground truth VIS images and suitable for auxiliary human visual recognition.

We failed to reproduce appearance of eyes and hair region because our pigment model holds only for skin region. This is however can be compensated by separating those region from skin region and applying different synthesis algorithm.

Although we only perform our spectral basis learning with frontal images with neutral expression, our synthesis method can handle probe image with different orientations and expressions with tolerable error.

6.2 Discussion and Future Works

Throughout the course of work, we have gained several insights from discussing our method for face synthesis. We aware of our requirement for use of special equipment to capture multi-band NIR images simultaneously could limits practical use of proposed method in real application. However, we envision that the ability of camera to capture multiple images at different wavelengths in NIR spectrum will eventually move into camera firmware, thereby making the multispectral NIR image acquisition easier.

For future work, few important issues still remain to be addressed. We list up our recommendation for future works as following, in no particular order.

- **Different illumination condition**
  Throughout our experiment, we have acquired our experimental dataset under controlled illumination condition. Thereby, we can determine the bias vector $\boldsymbol{b}$ for probe NIR images since it is approximately the same as the bias vector in training set. However, for real applications in future works, we need to consider the following cases: 1) illumination condition of the probe images acquisition is different from the training set, 2) illumination condition of samples in the training set are varied among themselves.

- **Camera with different spectral sensitivity**
  Scene radiance is recorded via camera sensors that specify light of different bandwidth to be observed. Therefore, observed signal values are dependent on spectral sensitivity of camera. For instance, same face can appeared to be slightly different in intensity when taken with different camera.

- **Ill-conditioned mixing matrix $\boldsymbol{R}_{nir}$**
  Throughout our experiments, we have encountered few trial experiments where the VIS images were not well synthesized. One of the reasons is the ill-conditioned $\boldsymbol{R}_{nir}$. We estimate pigment densities using least square estimation method, where we attempted to minimize the reconstruction error using Equation 4.5 in Section 4.4. When condition number of matrix $\boldsymbol{R}_{nir}$ is big (ill-conditioned), a small error in observed signal $\boldsymbol{r}_{nir}$ leads to greater error in pigment densities $\boldsymbol{q}$ estimated, and later causes the noise in synthesized VIS images. Ill-conditioned mixing matrix need to be handled so that the pseudo inverse matrix can be uniquely determined.

- **Band selection on NIR spectrum**
  Our work focused on the extraction of melanin and hemoglobin spectral basis from multispectral images and the use of the basis to synthesize VIS images. It is usual to think that performance of pigment extraction can be improved if we select specific wavelengths from NIR spectrum, which reflect optical properties of melanin and hemoglobin better. Heuristics on NIR band selection are needed to improve pigment spectral basis extraction.

References

[1] S. Li and A. Jain, *Encyclopedia of biometrics*, vol. 1. Springer Verlag, 2009.

[2] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[3] S. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 627–639, 2007.

[4] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[5] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Li, "Face matching between near infrared and visible light images," *Advances in Biometrics*, vol. 4642, pp. 523–530, 2007.

[6] B. Klare and A. Jain, "Heterogeneous face recognition: Matching nir to visible light images," in *Proc. IEEE International Conference on Pattern Recognition*, pp. 1513–1516, 2010.

[7] Z. Zhang, Y. Wang, Z. Zhang, and G. Zhang, "Face synthesis from near-infrared to visual light spectrum using quotient image and kernel-based multifactor analysis," in *Proc. IEEE International Conference on Multimedia & Expo*, pp. 1–4, 2011.

[8] J. Chen, D. Yi, J. Yang, G. Zhao, S. Li, and M. Pietikäinen, "Learning mappings for face synthesis from near infrared to visual light images," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 156–163, 2009.

[9] Z. Zhang, Y. Wang, and Z. Zhang, "Face synthesis from near-infrared to visual light via sparse representation," in *Proc. IEEE International Joint Conference on Biometrics*, pp. 1–6, 2011.

[10] M. Shao, Y. Wang, and Y. Wang, "A super-resolution based method to synthesize visual images from near infrared," in *Proc. IEEE International Conference on Image Processing (ICIP2009)*, pp. 2453–2456, 2009.

[11] K. S. Bersha, "Spectral imaging and analysis of human skin," Master's thesis, University of Eastern Finland, 2010.

[12] E. Angelopoulou, "Understanding the color of human skin," in *Proc. SPIE Conference on Human Vision and Electronic Imaging VI*, vol. 4299, pp. 243–251, 2001.

[13] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori, and Y. Miyake, "Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin," *ACM Transactions on Graphics*, vol. 22, pp. 770–779, 2003.

[14] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[15] Y. Suzuki, K. Yamamoto, K. Kato, M. Andoh, and S. Kojima, "Skin detection by near infrared multi-band for driver support system," in *Proc. 7th Asian Conference on Computer Vision-Volume Part II*, pp. 722–731, 2006.

[16] K. Sakashita, Y. Yagi, R. Sagawa, R. Furukawa, and H. Kawasaki, "A system for capturing textured 3d shapes based on one-shot grid pattern with multi-band camera and infrared projector," in *Proc. IEEE International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pp. 49–56, 2011.

[17] X. Tang and X. Wang, "Face sketch synthesis and recognition," in *Proc. 9th IEEE International Conference on Computer Vision*, pp. 687–694, IEEE, 2003.

[18] W. Konen, "Comparing facial line drawings with gray-level images: A case study on phantomas," in *Proc. International Conference on Artificial Neural Networks*, pp. 727–734, Springer, 1996.

[19] X. Tang and X. Wang, "Face photo recognition using sketch," in *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. I–257, 2002.

[20] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1005–1010, 2005.

[21] M. Reiter, R. Dormer, G. Langs, and H. Bischof, "3d and infrared face reconstruction from rgb data using canonical correlation analysis," in *Proc. 18th IEEE International Conference on Pattern Recognition*, vol. 1, pp. 425–428, 2006.

[22] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[23] A. Shashua and T. Riklin-Raviv, "The quotient image: Class-based re-rendering and recognition with varying illuminations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 129–139, 2001.

[24] M. Shao, Y. Wang, and P. Liu, "Face relighting based on multi-spectral quotient image and illumination tensorfaces," *Asian Conference on Computer Vision*, pp. 108–117, 2009.

[25] M. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *Proc. 7th European Conference on Computer Vision—ECCV 2002*, pp. 447–460, Springer, 2002.

[26] Y. Li, Y. Du, and X. Lin, "Kernel-based multifactor analysis for image synthesis and recognition," in *Proc. 10th IEEE International Conference on Computer Vision*, vol. 1, pp. 114–119, 2005.

[27] K. Jia and S. Gong, "Multi-modal tensor face for simultaneous super-resolution and recognition," in *Proc. 10th IEEE International Conference on Computer Vision*, vol. 2, pp. 1683–1690, 2005.

[28] H. Hariharan, A. Koschan, B. Abidi, A. Gribok, and M. Abidi, "Fusion of visible and infrared images using empirical mode decomposition to improve face recognition," in *Proc. IEEE International Conference on Image Processing*, pp. 2049–2052, 2006.

[29] A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, and D. Salesin, "Image analogies," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 327–340, 2001.

[30] R. Wang, J. Yang, D. Yi, and S. Li, "An analysis-by-synthesis method for heterogeneous face biometrics," *Advances in Biometrics*, pp. 319–326, 2009.

[31] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[32] A. Krishnaswamy and G. Baranoski, "A study on skin optics," tech. rep., CS-2004-01, School of Computer Science, University of Waterloo, Canada, 2004.

[33] J. Nordlund, R. Boissy, V. Hearing, R. King, W. Oetting, and J. Ortonne, *The pigmentary system*. Oxford University Press, 1998.

[34] S. Prahl, "Optical absorption of hemoglobin," *Oregon Medical Laser Center*, 1999. `http://omlc.ogi.edu/spectra/hemoglobin/index.html`.

[35] W. Zijlstra, A. Buursma, and W. Meeuwsen-Van der Roest, "Absorption spectra of human fetal and adult oxyhemoglobin, de-oxyhemoglobin, carboxyhemoglobin, and methemoglobin.," *Clinical chemistry*, vol. 37, no. 9, pp. 1633–1638, 1991.

[36] H. Nakai, Y. Manabe, and S. Inokuchi, "Simulation and analysis of spectral distributions of human skin," in *Proc. 14th International Conference on Pattern Recognition*, vol. 2, pp. 1065–1067, 1998.

[37] S. Jacque, "Skin optics," *Oregon Medical Laser Center News*, 1998. `http://omlc.ogi.edu/news/jan98/skinoptics.html`.

[38] R. Anderson and J. Parrish, "The optics of human skin," *Journal of Investigative Dermatology*, vol. 77, no. 1, pp. 13–19, 1981.

[39] M. Babaie-Zadeh and C. Jutten, "A general approach for mutual information minimization and its application to blind source separation," *Signal Processing*, vol. 85, no. 5, pp. 975–995, 2005.

[40] G. Zonios, A. Dimou, I. Bassukas, D. Galaris, A. Tsolakidis, and E. Kaxiras, "Melanin absorption spectroscopy: new method for noninvasive skin investigation and melanoma detection," *Journal of biomedical optics*, vol. 13, no. 1, p. 014017, 2008.

List of Publications

[1] ゴウキムシン，松川徹，岡部孝弘，佐藤洋一, “顔認識のための近赤外マルチバンド画像の可視光画像への変換,” 第 *15* 回画像の認識・理解シンポジウム *(MIRU)*, 2012.

[2] K.S. Goh, T. Matsukawa, T. Okabe, and Y. Sato, “Converting Near Infrared Facial Images to Visible Light Images using Skin Pigment Model,” *The 13th IAPR International Conference on Machine Vision Applications (MVA)*, May 2013. (Submitted)